Abstract

This article explores a typology of impersonal constructions [1] (constructions with arbitrary interpretations or arbs) in Russian, German, and English based on their semantic and discourse-pragmatic properties.

(1) They speak English in America.

The goal of this work is twofold. First, I want to introduce a meaning-based typological distinction into the realm of impersonals. I will argue that some arbs are semantically definite, while others are variable, drawing attention to previously unobserved behaviour of different arbs with respect to adverbial quantification. This divergent behaviour points to the inadequacy of any semantic analysis that treats the two types of arbs in the same fashion. I then point out further differences between the arbs (and non-arbitrary NPs) that emerge in their interaction with topic structure and discourse anaphora, which further corroborate the separation between the two types of arbs. Secondly, I want to use the study of arbs to explore the architecture of the semantics-pragmatics interface. I conclude that the behaviour of arbs forces a division of labour between semantics and pragmatics in the calculations pertaining to reference and referent-tracking.

Keywords
Arbitrary interpretation, Centering Theory, definite, variable, impersonal pronouns, Quantificational Variability Effect (QVE)

1 Introduction

This paper explores a typology of impersonal constructions in three European languages based on their semantic and discourse-pragmatic properties.

(1) They speak English in America.
I will refer to all of the constructions considered here by a descriptive cover-term constructions with arbitrary interpretations (arbs). Various authors have described as arbitrary the interpretations of pronouns and null syntactic elements (traces and PRO) that do not involve antecedents (Yule, 1982) or bound-variable interpretations (Jaeggli, 1986; Lebeaux, 1984; Cabredo-Hofherr, 2003, among others). These items then yield sentence interpretations that have a generic or impersonal flavour to them. Another property that the constructions considered here have in common is the demotion (in the several senses to be defined below) of the agent (cf. Myhill, 1997; Siewierska, 2008a). While these constructions have been extensively explored over the last several decades, the proposed analyses remain extremely diverse, and semantics of arbitrariness is far from being thoroughly understood. At the same time discourse-related properties of these constructions have received very little attention in the literature (Koenig and Mauner, 1999 and Prince, 2003, 2006 being two exceptions specifically focussing on discourse effects of arbs). Discourse effects of arbitrary pronouns are also mentioned in Egerland (2003) and related work (Cabredo-Hofherr, 2010, among others), who derive them from the featural composition of these pronouns.

The goal of this work is twofold. First, I want to introduce a meaning-based typological distinction into the realm of impersonals. I will argue that some arbs semantically definite, while others are variable, drawing attention to previously unobserved behaviour of different arbs with respect to adverbial quantification. This divergent behaviour points to the inadequacy of any semantic analysis that treats the two types of arbs in the same fashion. I then point out further differences between the arbs (and non-arbitrary NPs) that emerge in their interaction with topic structure and discourse anaphora, which further corroborate the separation between the two types of arbs. Secondly, I want to use the study of arbs to explore the architecture of the semantics-pragmatics interface. I conclude that the behaviour of arbs forces a division of labour between semantics and pragmatics in the calculations pertaining to reference and referent-tracking.

The arbs I focus on are the 3rd-person plural constructions with antecedentless pronouns in Russian and English (henceforth, 3PL ARBS), as well as the impersonal pronoun *man*.

---

1 I am extending here the usage of the term arbitrary from its application to the null subjects in the Spanish 3rd-person plural null pronouns [i] (compare with [ii]) (Suner, 1983), which in turn derives from its usage for the agents of certain infinitival clauses [ii] (e.g., Lebeaux, 1984).

i. Spanish (example and translation quoted from Cabredo-Hofherr, 2003)
   - Tocan a la puerta
     - ‘knock.3PL on the door’
     - ‘Someone’s knocking on the door’ (lit. ‘They’re knocking on the door’)

ii. [CP PRO To write a dissertation] is hard.

2 Note that my use of the term ‘arbitrary’ is different from the way it is used in some other work on impersonals, e.g., Egerland (2003). I term all of these constructions, rather than just one of their uses ‘arbitrary’.

3 These constructions are sufficient to demonstrate the interpretational distinctions between the two types of arbs. However, English, Russian, and German have a number of other constructions with arbitrary interpretations, including short verbal passives [i] in all three languages, *sja*-passives in Russian [ii], 3PL arbs...
3pl

3pl arbs

Russian

a. V Amerike govorjat po-anglijski.
‘In America speak.3PL in-English’
‘They speak English in America’

English

b. They speak English in America.

German impersonal pronoun *man* (from *Der Kleine Prinz*)

Man sieht nur mit dem Herzen gut.
‘MAN sees only with the heart well’
‘One can see well only with the heart.’

All three items under investigation have been claimed to have arbitrary or impersonal interpretations, in that they are “seen to have a subject but not a fully specified one, are typically identified with constructions in which a subject denotes a generic human or a loosely specified set of individuals” (Siewierska 2008a). These non-specific subjects are realised by pronouns; the constructions are “used when the intention of the speaker is to remain vague about the exact identity of the subject” (D’Alessandro (2004) on impersonal *si* in Italian).

Different types of impersonal constructions have been treated by a number of researchers in the last several decades (Malamud 2006; Jaeggli 1986; Cinque 1988; Condoravdi 1989; Kim 1991; Casielles 1993; Chierchia 1995b; Koenig and Mauner 1999; Alonso-Ovalle 2002; Egerland 2003; D’Alessandro and Alexiadou 2003; D’Alessandro 2004; Cabredo-Hofherr 2003, 2010; Prince 2003, 2006; Siewierska 2008a, 2010, 2011; Siewierska and Papastathi 2011; Hoekstra 2010, including), with different authors focussing on their syntactic and morphosyntactic properties, as well as various aspects of their semantics and pragmatics.

The semantics of arbitrariness in particular has received significant attention in the literature, with proposals for the interpretation of arbs ranging from uniformly indefinite analyses (e.g., Jaeggli 1986; Cinque 1988; Condoravdi 1989; Kim 1991; Chierchia 1995b), to work treating arbs as special kinds of definite pronouns (e.g. Alonso-Ovalle 2002; Hoekstra 2010), to accounts arguing that at least some of the arbs are ambiguous between several formal translations (Cabredo-Hofherr 2003), and finally proposals that treat arbs as a category in German, English impersonal pronoun *one iii*, and 2nd-person impersonals [iv] in all three languages. I will address English and Russian passives and Russian *sja*-passives in the discussion in §4.

---

i. The striking shopmen were blacklisted...

ii. P’jer i Marija mogut schitat’sja primerom... beskorystnogo sluzhenija nauke.

Pierre and Maria can consider.sja example.INSTR selfless.GEN service.GEN science.DAT

‘Pierre and Maria can be considered an example of selfless service to science.’

Russian-edition preface to *Madame Curie*

iii. One sees well only with the heart.

iv You never know whatcha gonna get

Valley of the Moon

The Little Prince

Forrest Gump
of their own that cannot be assimilated to either definite or indefinite non-arbitrary NPs (Koenig and Mauner 1999). In this work I will be focussing on one specific aspect of semantic (in)definiteness — definite and indefinite NPs are distinguished in their sensitivity to adverbial quantification, and arguing that the arbs in [2], like non-arbitrary definite NPs, are insensitive to such quantification, and while those in [3] are sensitive to it, patterning with indefinite NPs.

I will then show that, at least for these arbs, these semantic properties correlate with pragmatic ones. Existence of two different types of arbs, with different effects on reference and coherence of subsequent discourse, serves as evidence against prior proposals for the pragmatics of arbs in Koenig (1999); Koenig and Mauner (1999); Egerland (2003). In addition, this investigation of the discourse effects of various arbs points to a salience/topicality scale among arbitrary and non-arbitrary NPs, with fine-grained distinctions that cannot be captured by the semantic (in)definiteness dichotomy. These distinctions require a separation between categorical semantic calculations, and gradient pragmatic ones.

The structure for the remainder of the paper is as follows. In the rest of this section, I introduce the arbitrary constructions to be considered, including their distribution, and some of their morphosyntactic and interpretive properties. At the end of the section, I introduce the typology to be defended in the rest of the paper, and place this proposal in the context of formal semantics and pragmatics of referential expressions. Section 2 discusses the semantic behaviour of the arbs, and provides evidence for a semantic distinction between two types of arbs that makes them in some ways parallel to non-arbitrary NPs. In this section, I also discuss the consequences of the typology for the semantic analysis of these constructions. Section 3 provides further corroborating evidence for the two types of arbs based on their effect on subsequent discourse, while also demonstrating surprising differences between, on the one hand, the definite and variable arbs, and on the other hand, the non-arbitrary definites and indefinites. Section 3 includes the classification of other arbs in the present typology, specifically addressing similarities and differences between the impersonal pronouns of both types and implicit agents in passives. Finally, section 5 provides concluding discussion, focussing on the consequences this study of arbs has for the architecture of the grammar, and specifically the semantics-pragmatics interface.

1.1 The distribution of 3pl arbs and German man

The constructions I consider — 3pl arbs, and the dedicated German impersonal pronoun man — are pronominal subjects which do not have antecedents in prior discourse (Siewierska 2008a). They fit the semantic definition of impersonality which is based on reference: “constructions which are considered to be impersonal in this [...] sense of the term depict situations and events which may be brought about by a human agent, but crucially one which is not specified” (Siewierska 2008b). Let me introduce and exemplify these constructions now.
1.1.1 The uses of 3pl arbs in Russian and English

Antecedentless third-person plural pronouns — the 3pl arbs — have been studied in a number of European and non-European languages (cf. Siewierska and Papastathi (2011) and references therein). Here, I will consider 3pl arbs in Russian [2a] and English [2b].

(2) 3pl arbs

Russian

a. V Amerike govorjat po-anglijski.
   ‘In America speak.3pl in-English’
   ‘They speak English in America’

English

b. They speak English in America.

The constructions with 3pl arbs in both languages are said to always exclude the speaker. However, in some examples this exclusion seems to be a matter of perspective rather than reference. For instance, an American can utter [2b] to a hearer who doesn’t know what languages are in use in the US. In Russian as well, an example like [4] can be used by a chef explaining her own recipe.


Ulitok kladut v sol’ na celyj den’ i gotov’at s chesnokom
Snails.ACC put.3pl in salt for whole day and prepare.3pl with garlic.

‘One puts snails in salt for a whole day and prepares them with garlic.’

The translational equivalents in [2a, b] have a near-universal interpretation, paraphraseable as “People in America speak English.” This is the quasi-universal reading so-called by Cinque (1988).

Note that the Russian sentences in [2a, 4] have no overt subject — the 3rd-person plural agreement on the verb signals the morphosyntactic nature of the silent subject. This is surprising, because non-arbitrary 3rd-person pronouns in Russian are generally overt. In the sense that most of its pronominal subjects are overt, Russian is not a pro-drop language, as noted, among others, in Franks (1995).

Unlike non-arbitrary pronouns, the 3pl arbs in subject position are obligatorily null in Russian [3] (corpus search provides several examples that may be overt 3rd person plural impersonals; however, these are extremely rare).

4The claims regarding the interpretation of 3pl arbs that I make here only apply to this construction in these two languages. Many previous studies addressed 3pl arbs in a variety of languages, e.g., Spanish (Suñer, 1983; Alonso-Ovalle, 2002) and Greek (Condoravdi, 1989). However, properties of arbs, including 3pl arbs, vary both cross-linguistically (Siewierska (2010), cf. the discussion in Malamud (2012a)), and diachronically (Siewierska (2010), among others). Thus, it is quite possible that an arb would vary with respect to the semantic and pragmatic properties explored in the present study.

Rizzi (1986) proposes four types of languages with respect to pro-drop:

• languages where all subjects must be overt
• languages where only expletive subjects are null
• “partial pro-drop” languages with quasi-argument pro, where subjects of weather predicates as well
This provides a means of distinguishing 3pl arbs (always null) from the discourse-anaphoric 3rd-person plural pronouns (almost always overt) in Russian — a formal distinction unavailable in English, where personal and impersonal uses of they are both overt, or in pro-drop languages, which drop their personal subject pronouns most of the time. Researchers argued for a connection between the distribution of uses of 3pl arbs and the pro-drop properties of the language (Cabredo-Hofherr, 2006; Siewierska, 2011).

What are these uses? Building on much prior work, Cabredo-Hofherr (2003) proposes five distinct types of interpretations for 3pl arbs. The uses are listed in (6) below, together with labels that Cabredo-Hofherr (2003) gives to them.

(6) Cabredo-Hofherr's (2003) readings of 3pl arbs

<table>
<thead>
<tr>
<th>Use</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. V dver’ stuchat!</td>
<td>Specific existential</td>
</tr>
<tr>
<td>‘Someone is knocking on the door!’</td>
<td></td>
</tr>
<tr>
<td>b. They found a motorbike in the courtyard.</td>
<td>Vague existential</td>
</tr>
<tr>
<td>c. Zdes’ rybu zavorachivali.</td>
<td>Inferred existential</td>
</tr>
<tr>
<td>‘Here fish was wrapped’ (pointing at a stain on paper)</td>
<td></td>
</tr>
<tr>
<td>d. They’ve raised the taxes again.</td>
<td>Corporate</td>
</tr>
<tr>
<td>e. They speak Spanish in Spain.</td>
<td>Universal</td>
</tr>
</tbody>
</table>

In an empirical investigation of 3pl arbs, Siewierska and Papastathi (2011) propose a sixth use with verbs of saying (they say), arguing that this use must be distinguished from others, because there are languages in which only this use of 3pl arbs is possible. Previous researchers have treated this use as a subclass of vague 3pl arbs.

The readings where the arb is paraphrasable as someone have been called existential (Jaeggli, 1986; Cabredo-Hofherr, 2003) and quasi-existential (Cinque, 1988). The specific existential use, such as [6a], describes an event anchored at a particular point in time, and

<table>
<thead>
<tr>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>“full pro-drop” languages, which in addition to quasi-argument and expletive pro also allow referential pro, like Italian and Spanish where the majority of pronominal subjects are null.</td>
</tr>
</tbody>
</table>

Note, however, that the non-overtness of the 3pl arbs in Italian and Spanish is not a tendency, but a grammatical requirement (cf. Jaeggli, 1986 for Spanish). Cabredo-Hofherr (2006) modifies and refines this proposal, further separating deictic pro, present in languages like Finnish that permit 1st and 2nd person pronouns to be dropped, and distinguishing anaphoric 3rd person pro from non-anaphoric, arbitrary 3rd person pro.

In Russian, expletives, subjects of weather predicates, and arbs are obligatorily null, while referential subject pronouns are overt. Thus, Russian is not a full pro-drop language, but a partial pro-drop one, and drops its expletives, quasi-arguments, and arbs, but not other subjects.
the arb refers to a specific person(s) that the speaker has some contact with. Vague existential
uses of 3PL arbs, illustrated in [6b], simply assert that an event has taken place, without
referring to a specific time. Inferred existentials such as [6c] are very similar to the vague
uses, but there is a result state that allows the speaker to infer the existence of the event.
The properties of the inferred existential arbs, thus, are closely linked to the properties of
the perfect predicate with which these arbs are used.

Cabredo-Hofherr argues for a separation between the three existential readings based on
cross-linguistic evidence. First, [6a, c] are ruled out in French 3PL arbs, thus separating these
readings from the “vague existential”. At the same time, she states that [6c] but not [6a] are
ruled out in Moroccan and Syrian Arabic. A native speaker survey study in Siewierska and
Papastathi (2011) shows that specific and inferred uses are at best marginal in English; this
is confirmed by my data. For instance, note that the specific example in [6a] is not possible
with English 3PL arbs. I collected evidence from corpus data of spoken and written Russian
as well as a native speaker survey, which definitively points to these readings being perfectly
acceptable in Russian.

Defining properties of the existential uses [6a-c] are debated: while earlier investigations
of 3PL arbs focussed on episodic sentences (Jaeggli 1986; Cinque, 1988), studies by Con-

doravdi (1989); Alonso-Ovalle (2000) point out that existential uses are also available in
characterising sentences. They give examples such as [7] below, where a locative individual-
level expression is present, making these cases similar to the (quasi)-universal cases in [2,6e].

(7) They sell ice-cream on the corner (based on Condoravdi (1989))

As pointed out in Siewierska and Papastathi (2011), it is extremely difficult to draw a
strict line between the vague existential and the corporate use. Thus, the following charac-
terising example can be thought of as either vague ‘Someone keeps renaming it’ or else as
corporate ‘The owners keep renaming it’.

(8) This hotel used to be called ‘Lux’. They keep renaming it though.

The corporate use of 3PL arbs [6d] is associated with predicates that have a culturally
designated subject, e.g. it is people in the government who raise taxes (Pesetsky 1995). The
survey study in Siewierska and Papastathi (2011) suggests that the vague use is “dispreferred
to the corporate” in English; my data shows that this is not so in Russian. The nature of
the difference between vague and corporate uses in Russian and English calls for a careful
investigation, which is outside the scope of this paper.

The sentences in [2,6c] exemplify the use of 3PL arbs that various authors have termed
universal (Cabredo-Hofherr 2003), quasi-universal (Cinque 1988), and generic (Siewierska,
2011) among others).

This last term seems to suggest that these readings of 3PL arbs are only available in
generic contexts. However, this is not the case — the quasi-universal interpretation of 3PL
arbs is independent from aspectual specification of the clause, as has been pointed out by
Alonso-Ovalle (2000) ⁶. For instance, in [9], the 3PL arb is interpreted as “(most) people in

⁶Aspectual specification of the clause, or other properties of the temporal or event structure may come
into play in other languages. For example, as an anonymous reviewer points out, in French the following
France” despite the episodic context.

(9) Russian (adapted from Cabredo-Hofherr (2003))

Vo Francii sejchas sp’at.
In France now sleep.3PL

‘In France, they’re sleeping now.’

However, there is a restriction on this use: Alonso-Ovalle (2000) points out, following Casielles (1993), that “the quasi-universal readings of these ... arbitrary pronouns are dependent on the presence of some kind of overt restrictor that appears under the form of a pseudo-locative expression,” such as in Spain in [2, 6e], or in France in [9].

The corporate [6f] and universal [6e] readings are claimed to not be paraphraseable using the indefinite someone. Note, however, that this is a property of the specific examples, rather than of the use of 3PL arbs with designated-subject predicates and locative-licensed sentences. For instance, sentences in [10] are paraphraseable using someone, as long as information from the designated subject and from the locative is retained [11].

(10) a. They took her to the Lubjanka prison yesterday. (adapted from Gulag Archipelago)
    b. They speak Spanish in this store.

(11) a. Someone from the KGB took her to the Lubjanka prison yesterday.
    b. Someone in this store speaks Spanish.

Another property groups the corporate uses together with existential, rather than universal ones. Jaeggli (1986); Cinque (1988) observe that existential readings of 3PL arbs in Spanish are limited to sentences where these arbs are non-derived subjects — that is, to transitive and unergative verbs. In fact, as noted, among others, in Siewierska (2011), the specific, vague, inferred and corporate uses of 3PL arbs [6a-d] are limited to such agentive verbs cross-linguistically.

Siewierska and Papastathi (2011) investigate the distribution of the five readings [6] in several European Languages. Their data gives rise to the following typological hierarchy, where a language in which a reading further to the right is acceptable will also have the uses to the left.

contrast obtains between the characterising sentence in [i] and the episodic sentence in [ii].

i. En Amérique, ils parlent / on parle anglais.
    ‘In America, they speak.3PL / ON speak.3s English.’

    ‘They speak English in the US.’

ii. En France, ils dormant / *on dort en ce moment.
    ‘In France, they sleep.3PL / ON sleep.3s in this moment.’

    ‘In France, they are sleeping right now.’

I am grateful to an anonymous reviewer for pointing out these examples. See also D’Alessandro (2004); D’Alessandro and Alexiadou (2003) for similar phenomena in Italian.
Implicational hierarchy of 3PL arb uses (Siewierska and Papastathi (2011), [28])
universal/corporate > vague > inferred/specific

Next, turning to the sixth speech-act use of 3PL arbs, Siewierska and Papastathi (2011) classify only the generic sentences with such arbs as a separate use, categorising episodic clauses as either vague or corporate. Like the vague readings, the speech act uses of 3PL in English are more limited than in Russian, as the contrast in [13] shows.

(13) Russian (from RNC)
   a. Chasto hval’at morskoj vozduh
      ‘Often praise.3PL sea air’
      ‘People often praise sea air’
   b. Nas chasto sprashivajut ‘Zachem vas tak mnogo na scene’
      ‘We.ACC often ask.3PL ‘Why you.GEN so many on stage’
      ‘We’re often asked, ‘Why are there so many of you on stage?’’
   c. Prozvuchal fragment sed’moj simfonii Shostakovicha,
      ‘Sounded fragment seventh.GEN symphony.GEN Shostakovich.GEN,
      kotoruju nazyvajut ‘Leningradskoj’
      which.ACC call.3PL ‘Leningrad’
      ‘We just played a fragment of the seventh symphony of Shostakovich, which is called ‘the Leningrad.’

English — judgements on arbitrary interpretation (all ok on non-arbitrary reading)
   a’. # They often praise the sea air.
   b’. # They often ask us ‘Why are there so many of you on stage?’
   c’. ? We just played a fragment of the seventh symphony of Shostakovich, which they call ‘the Leningrad’.

I will leave the cross-linguistic distinctions in the speech-act use of 3PL arbs to future research – since these are closely tied to specific verbs, it is likely that such distinctions are dictated by the selectional properties of these verbs.

Finally, a language like Russian, in which the lexicalisation of arbitrary and referential pronouns is different in vast majority of cases, allows us to distinguish a potentially new use of 3PL arbs. Specifically, the pronouns in the second sentence in [14a, b] can be interpreted as anaphoric to the arb in the first sentence. It has been observed that this kind of anaphoric use is not subject to the same restrictions as the initial arb (as observed in Cabredo-Hofherr (2006) for Spanish).

(14) Russian
   a. Govor’at, v Nicce horosh to.odyhat’. Chasto hval’at morskoj vozduh...
      ‘Say.3PL in Nice good to. vacation Often praise.3PL sea air’
      ‘They say it’s good to vacation in Nice. They often praise the sea air’
b. Govor’at, v Nice horosho otvyhat’. Oni chasto hval’at morskoj vovduh...
‘Say.3PL in Nice good to.vacation They often praise.3PL sea air’
‘They say it’s good to vacation in Nice. They often praise the sea air.’

The interpretation of the second-occurrence impersonal in [14a] resembles that of the definite personal pronoun in [14b]. However, there is an important difference between the two: while the definite oni in [14b] must, in the absence of other context, be co-referential with the preceding arb, in the impersonal [14a] the two arbs may also be referring to different (groups of) people, since the second arb does not have to be anaphoric.

In sum, six uses of 3PL arbs have been proposed in the literature; I identify the anaphoric use as a potentially new one to consider. The availability of the uses varies cross-linguistically; the distribution of the available uses within a language can also vary. Can these uses be analysed in a uniform manner, using a single denotation for the arb, and deriving the various readings through its interaction with contextual factors? In §2 I suggest that they can, contra previous proposals by Alonso-Ovalle (2002); Cabredo-Hofherr (2003), but first let us turn to an overview of German man.

1.1.2 The uses of German man

The dedicated impersonal pronoun man in German is the subject of sentences it occurs in, always triggering 3rd-person singular agreement on the verb [3, 15].

The distribution of this arb is in many ways similar to that of the 3PL arbs introduced above, though there are important differences. Egerland (2003) distinguishes three readings for man-type impersonals in Swedish and French: a use termed generic [3, 15a], a use called arbitrary or existential [15b], and a use termed specific or inclusive [15c]; subsequent researchers adopt them for the German man (Cabredo-Hofherr (2004, 2010); Siewierska (2011), among others).

(15) a. Man soll sich zweimal am Tag die Zähne putzen.
‘One should brush teeth twice a day.’ (from Cabredo-Hofherr 2010)

b. Gestern hat man mir mein Rad geklaut.
‘Yesterday, someone stole my bike.’ (from Cabredo-Hofherr 2010)

c. Als ich klein war, wurde man nur am Freitag gebadet.
‘When I was little, we only had a bath on Fridays.’ (from Kratzer 1997)

According to Egerland (2003), the generic reading is available in characterising, generic sentences; while the arbitrary is restricted to episodic contexts with specific time reference. Thus, Egerland (2003) and others claim a correlation between the aspectual specification of
the clause and the available readings of these pronouns. In fact, the correlation with aspect is imperfect, since generic readings are sometimes available in seemingly episodic contexts.

(16) Gestern feierte man in Baden-Württemberg “Heilige Drei Könige”.
   ‘Yesterday celebrated MAN in Baden-Württemberg “Holy Three Kings”
   ‘People in Baden-Württemberg celebrated the holiday “Heilige Drei Könige” yesterday.’

Conversely, there are examples in which man receives existential interpretation in generic contexts.

(17) German
   Da spricht man Spanisch.
   ‘There speaks MAN Spanish’
   ‘Spanish is spoken there.’

Like many of the uses of 3pl arbs, existential uses of German man are limited to non-derived, agentive, subjects.

The constructions I am investigating in this paper vary with respect to their exclusion or inclusion of the speaker and hearer. Specifically, unlike 3pl arbs, man has speaker-inclusive readings in most varieties of German. However, as argued in Zifonun (2000) and Malamud (2012a), among others, this speaker-inclusion is not of the same nature as the first-person reference in personal pronouns. Elsewhere, I conduct a thorough investigation of the nature of the inclusive reading and its relationship with other uses of man (?). Here, note that even on this reading, man is not always synonymous with the definite first-person pronouns wir ‘we’ or ich ‘I’. For instance, unlike wir, inclusive man cannot denote humankind (cf. the discussion of kinds in §2 below), and cannot enter into antecedence relations with 1st person pronouns.

The factors that affect availability of the inclusive reading for German man are subject to enormous inter-dialectal variation; I shall not describe or attempt to explain them here.

Like 3pl arbs, sequences of sentences with man can create an impression of co-reference between the different occurrences of the pronoun. In addition, there is another use of man:

i. Se abrió las ventanas en dos segundos.
   ‘They opened the windows in two seconds.’

---

7 This relationship with aspect is not uniform across the range of arbs for which such a connection has been claimed. For instance, D’Alessandro and Alexiadou (2003) provide a detailed discussion of the inclusive reading of Italian si impersonals. They claim that in sentences with perfective aspect, only the inclusive reading is available. This generalisation does not apply to German man, where the existential (non-inclusive) reading is available in sentences like [16b]. The difference is not due to the differing origins (a common noun meaning person vs. reflexivity marker) of these arbs, since Zagona (1999) gives the following non-inclusive example for Spanish se, which is a cognate of Italian si:

   i. Se abrió las ventanas en dos segundos.
      ‘They opened the windows in two seconds.’ (translation as given in Zagona 1999)
that relates to a previous occurrence of the same pronoun in quantificational constructions. On this use, the two occurrences of man resemble bound variables co-varying under quantification [18].

(18) German

Wenn man klug ist, ist man gewöhnlich stolz.
‘If man smart is, is man usually proud’

‘If a person is smart, he/she is usually proud.’

In §2 I will discuss the uses corresponding to the second occurrence of man.

In sum, the 3pl arbs in English and Russian and the German man can have a variety of readings, which are not directly correlated with the aspectual specification of the clause in which they occur. Quasi-universal readings of 3pl arbs are licensed by the presence of a locative in both episodic and characterising sentences, while generic readings of man do not need such licensing in characterising contexts. The uses of 3pl arbs that are not licensed by a locative, and existential uses of man are restricted to non-derived subjects. In addition, while generalisations expressed using 3pl arbs can include the speaker, an examination of episodic sentences shows that these arbs cannot directly refer to the speaker, or a speaker-inclusive plurality, in contrast to speaker-inclusive uses of man.

Speaker-inclusion aside, the ranges of interpretation available for these constructions may seem very similar. However, a closer examination of their semantic and pragmatic behaviour reveals important differences. We turn to these next.

1.2 The typology

The chief claim I will be defending here is given in [19] below:

(19) Arbs fall into two types that differ in their semantic behaviour and their effects on subsequent discourse. The 3pl arbs belong to the first type, which I will term DEFINITE-TYPE, while German impersonal pronoun man exemplifies the second type, which I will term VARIABLE-TYPE.

a. Definite-type arbs behave like (plural) definites, while variable-type arbs pattern with semantically indefinite-like variables.

b. Definite-type arbs are possible, but unlikely (rare) topics and antecedents for discourse anaphora, while variable-type arbs do not participate in topic structure or discourse anaphora at all.

In the context of non-arbitrary referential NPs, this proposal contains a tension between the semantic properties of arbs in [19a] and their pragmatic properties in [19b]. Formal semantic theories have drawn a connection between discourse anaphoric potential of indef-
Theory (Kamp, 1981; Kamp and Reyle, 1993), indefinites introduce new discourse referents, which enables them to antecede pronouns, and simultaneously makes them susceptible to quantification. Thus, it is unexpected that arbs that pattern with indefinites with respect to quantification (variable-type arbs like German *man*) are discourse-anaphorically impotent.

In addition, [19b] points to a threefold pragmatic distinction between non-arbitrary definites and indefinites, definite arbs, and variable arbs, which does not correspond to the two-way semantic definiteness distinction. The effects of arbs on their linguistic and extra-linguistic context has received relatively little attention in previous literature, with existing discussion focussing on their referential properties. Authors concerned with syntactic and semantic properties of German *man* or its Yiddish equivalent, for instance, noted that while it can support reflexive anaphors, it is unable to provide antecedents for intersentential anaphora (Kratzer 1997; Prince 2003, 2006; Zifonun 2000; Cabredo-Hofherr 2010; Malamud 2006). In fact, the only item that can be used to refer to the denotation of *man* (outside of very local contexts licensing reflexive anaphors) is *man* itself. In contrast to *man*, 3PL arbs can support intersentential anaphora — a fact that emerges most clearly in Russian, where personal and arbitrary 3rd-person plural pronouns have different realisations (overt and null, respectively). A corpus study of the discourse consequences of the use of *man* and of 3PL arbs is presented in §3.

Koenig (1999) and subsequent work (Koenig and Mauner, 1999; Koenig et al., 2002) argue that discourse-inert arbs like the French dedicated impersonal pronoun *on* are similar to implicit agents in short verbal passives. In §3 I shall show that 3PL arbs are not referentially impotent, in contrast to dedicated impersonals like *on* and *man*. This contrast is evidence against Koenig’s uniform analysis of the pragmatic behaviour of arbs. In §5 I build on Prince (2006) in arguing that a purely semantic framework is inadequate to capture the discourse effects of arbs and non-arbitrary NPs. This has important consequences for the division of explanatory labour between semantics and pragmatics.

Siewierska (2011) describes a distributional distinction between *man*-like arbs and 3PL arbs and its relation to pro-drop in the languages of Europe, building on Cabredo-Hofherr (2010). Arbs originating in the noun *person*, such as *man*, are restricted to generic contexts in pro-drop languages, but can appear in episodic context in non-pro-drop languages. English *one* is claimed to be a *man* type impersonal by, eg, Safr (2004): “The Germanic form for the generic pronoun, modern German *man*, old English *man*, and middle English *me*, appears to have been replaced by French *on*, which derives independently from a similar root...” If this etymology is correct and English *one* is a *man*-type impersonal, then it presents an exception to the trend described by Siewierska, because it is restricted to generic environments. In contrast, 3PL arbs can appear in both episodic and generic contexts in pro-drop languages, while in non-pro-drop languages tend to be restricted to generic uses. Indeed, 3PL arbs can appear in episodic contexts in Russian, which is a partial pro-drop language (cf. footnote 5), but not in English, which is not pro-drop. However, it is not clear that Russian fits Siewierska’s definition of pro-drop. For instance, she notes that “3PL-IMPS... are favoured over MAN-IMPS in pro-drop languages by virtue of the overall preference for bound pronominal subjects that these languages display.” This preference clearly does not hold in Russian, where the majority of subject pronouns are overt free morphemes, in stark contrast to full pro-drop languages like Italian.

The findings of Siewierska’s study and the present investigation are compatible: if 3PL arbs and *man* have different semantic and pragmatic effects, it is unsurprising that their distribution is different. The connection to the pro-drop parameter is interesting, but since it is a trend, rather than a categorical rule, it should not be derived from the denotation of arbs.
Let us now turn to the empirical investigation of the semantic and pragmatic behaviour of arbs. In the next two sections I present evidence for the semantic \[19a\] and pragmatic \[19b\] parts of the typology.

2 Semantics of arbs

In this section, I first establish a test that can detect semantic (in)definiteness in arbs, in one specific sense of indefiniteness. In particular, the test can determine whether an arb is susceptible to adverbial quantification. Subsequently, the test demonstrates that 3rd-person plural arbs are not susceptible to such quantification, while German man is. This indicates a fundamental difference between the semantics of these arbs. I discuss other empirical clues to this semantics, and propose a definite analysis for 3pl arbs and an indefinite-like analysis for man.

2.1 The QVE test, definites, and indefinites

Lewis (1975) notes that quantificational adverbs (Q-adverbs) seem to quantify over the denotation of indefinites (Quantificational Variability Effect, henceforth QVE). Q-adverbs denote quantifiers that at least sometimes target situation or event variables: always (for every situation/event), usually (for most situations/events), etc. (see Lewis (1975); Berman (1987); de Swart (1991); von Fintel (1995); Chierchia (1995a) for influential analyses of Q-adverbs). These adverbs yield the QVE in English \[20\] with singular garden-variety indefinites and bare plurals \[10\], both in single-clause sentences \[20a,b\], and in ‘donkey’ \[11\] conditionals \[20c\]. With the right context, definites can yield the effect of varying, too \[20d\].

(20) Quantificational Variability Effect (QVE) in English

(QVE available: Most students = QVE on students)

a. A Penn student is usually/rarely smart.

b. Penn students are usually/rarely smart.

c. If a student here deals with the Mafia, he always/usually/sometimes gets killed.

d. In the admissions process, we interview prospective students by inviting them into the room one at a time. If the kid is tall, he is usually smart.

\[9\]Of course, there is always an option of Q-adverbs quantifying over times; I will ignore these temporal readings here, since they tell us nothing about the semantics of the nominals involved.

\[10\]As an anonymous reviewer points out, English bare plurals are kind-denoting, and have been argued to be definite from a semantic standpoint (Carlson 1977 and subsequent work). In German, kind-denoting plurals are expressed using a definite article. Kind plurals can undergo QVE cross-linguistically, as I discuss below. Thus, the QVE can eliminate the possibility of a non-kind definite interpretation; to check whether arbs are kind-denoting I will use a different test.

\[11\]The ‘donkey sentences’ (after famous examples due to Peter T. Geach 1962) are those that involve a pronoun which is interpreted as being bound by an operator, yet which is located outside the syntactic scope of that operator.

\[12\]I am grateful to three anonymous reviewers for providing this example and two others just like it.
All sentences in [20] have the effect of the adverb quantifying over the students, quantificational force of the NP varying with the adverb. Thus, [20, a, b] have the QVE reading Most/Few Penn students are smart, and [20c] has the QVE reading Most/Few of the student dealing with the Mafia get killed.13

To give an explicit semantic analysis of QVE, I use Dynamic Montague Grammar (DMG) (Groenendijk and Stokhof [1990], Chierchia [1995a], Honcoop [1998]), and specifically the version presented in Chierchia [1995a].14 In DMG, indefinites introduce dynamic existential quantification (the symbol for the dynamic existential quantifier is \( \exists \)), capable of binding pronouns in subsequent discourse. Q-adverbs are capable of binding individuals or events directly — this maintains compositionality and uniform treatment of quantification, while capturing Lewis’s (1975) insight that Q-adverbs are unselective quantifiers. The adverb creates a tri-partite quantificational structure, as illustrated in [21, 22], consisting of the operator (Q-adverb), a restrictor which can be articulated by an if-clause of a conditional, such as if a kid is tall in [21], and a nuclear scope, such as he is smart in [21].

The operation of Existential Disclosure (Dekker [1993], Chierchia [1995a]) can make indefinite NPs available for adverbial quantification by turning them into properties that restrict the Q-adverb [21].

(21) Usually, [if a kid is tall], [he is smart]

13A note on aspect and tense of QVE examples. I avoid stage-level predicates, because they highlight the temporal reading of Q-adverbs, making it difficult to detect the QVE readings. The “longer-term” stage-level and individual-level predicates (be smart, love someone), which tend to be stative, cause infelicity in the absence of QVE — e.g., the temporal reading of be smart indicates fluctuation of intelligence over time.

However, many achievements or accomplishments are also fine in testing for QVE, because they are totally incompatible with temporal readings of Q-adverbs. The individual-level readings for those aspectual types are usually “result-state” readings, as illustrated in [i].

i. In this club, a climber usually has climbed Mt. Everest
   (QVE available: most climbers in this club are in the state of having climbed Mt. Everest).

Finally, while most of my examples are in the present tense, past tense examples work in exactly the same way, as in [ii] below. All QVE sentences were tested with native speakers in both present and past tense, with identical results.

ii. a. A medieval university student was usually smart. (QVE available)
    b. The students at Wittenberg in 1517 were usually smart. (No QVE: only the temporal reading: now they’re smart, now they’re not)

14Influential treatments of QVE include LF-based approaches in Heim [1982], Diesing [1992], and the Discourse-Representation Theory (DRT) in Kamp [1981], Kamp and Reyle [1993]. These accounts propose that both definites and indefinites are referential, and blame their different behaviour on their referents: indefinites introduce new variables which Q-adverbs quantify over, while definites pick up old variables and are unavailable for quantification.


Unlike DRT, DMG provides an easy way to distinguish two kinds of indefinites — quantificational ones, and bare variables — which behave identically with Q-adverbs, but differ in their scope properties.
Thus, in (21a), the quantified variable \( x \) (to which the predicate \( \text{kid} \) applies) is disclosed by equating it with a new variable \( u \) over which lambda-abstraction takes place. This operation creates a predicate \( \text{tall kid} \) (21b,c), which can serve as the restrictor for the adverbial quantifier \( \text{usually} \) (translated as “most”’), just as the predicate \( \text{smart} \) serves as the nuclear scope for this quantifier.

In contrast, a definite in (22) cannot be turned into such a property without the presence of additional variables that might be disclosed, as in (20d). This is because Existential Disclosure can only reach a variable that is free or bound by an externally dynamic operator.

The effect is similar for plural definites and indefinites. Chierchia (1995a) includes a relation to context in his denotation for definite NPs (22), \( R \) is a free variable), similar to the situation-dependent treatment of definites in situation semantics (but see Schwarz (2009) for a discussion of important differences). In cases such as (20d), the variable over admission interviews provides the link from the definite NP to the external context. The Q-adverb can only apply to the variable representing the contextual parameter in (22), since the variable to which \( \text{kid} \) applies in (22) is not available for disclosure, unlike the existentially quantified variable in (21). Thus, in contexts that do not make available such variables over admission interviews, no QVE is predicted with the definite NPs (22).

Kind-denoting NPs with or without definite determiners yield QVE with Q-adverbs (23a, b) (As pointed out by Hoekstra (2010) for West Fresian).

(23) English

a. Lions are always/rarely/usually intelligent.

(QVE available: most/few/some lions = QVE on lions)

b. The African lion is always/rarely/usually intelligent.

(QVE available: most/few/some lions = QVE on lions)

In contrast to indefinites and kinds, QVE over a plural non-kind definite that does not come equipped with disclosable contextual variables is achieved with operators like \textit{for the most part} (24), not with Q-adverbs (25) (Nakanishi and Romero, 2004).
The students sitting over there now are for the most part smart.
(QVE available: Most students = QVE on students)

# The students sitting over there now are usually smart.
(QVE not available: the only reading is ‘Now they are smart, now they are not’)

The same pattern holds in Russian and German, ceteris paribus. Q-adverbs generate QVE with Russian and German indefinites:

the same adverbs fail to create variability with non-kind definites.

Instead, quantification over parts must be used to create the QVE. Like their single-clause counterparts, donkey conditionals in Russian and German function exactly as in English.

In German, kind-denoting plurals are formally definite, rather than bare as in English and Russian (Chierchia, 1998). They yield QVE as in [26], which patterns like the English bare plural in [23a], and unlike German non-kind definites in [27].

(26) German kinds (QVE: most/few/some lions)
Die Löwen sind normalerweise / selten / manchmal intelligent.
‘The lions are usually / rarely / sometimes smart’

‘Lions of this kind [e.g. African lions, zoo lions] are usually/rarely/sometimes smart.’

(27) German non-kind definites (no QVE)
# Diese Löwen sind normalerweise / selten / manchmal intelligent.
‘These lions are usually / rarely / sometimes smart’

‘These lions are usually/rarely/sometimes smart.’

Thus, susceptibility to QVE with Q-adverbs can serve as an empirical test for (in)definiteness in English, Russian, and German, as long as kind terms are detected using a different test.

The different behaviour of definites and indefinites with respect to QVE lies at the heart of the semantic (in)definiteness distinction, and is reified in practically every formal semantic framework, as exemplified for DMG in [21, 22]. I will therefore rely on the QVE test to provide an initial separation of the definite and indefinite-like arbs.

2.2 3PL arbs are definite.

Many researchers have claimed that the 3PL arb is uniformly indefinite (Jaeggli (1986); Condoravdi (1989); Kim (1991), Chierchia (1995b, p. 547), among others), or is ambiguous between definite and indefinite denotation (Cabredo-Hofherr 2003). In this section, I provide evidence that an indefinite translation for 3PL arbs is, at least in some contexts, unavailable.

2.2.1 3PL arbs are not susceptible to QVE with Q-adverbs

Malamud (2004, 2006) notes that QVE readings with 3PL arbs in Russian and English are not achieved with Q-adverbs [28].

Note that all examples in [28] include a locative. However, under any account, this should not force a definite reading.
Instead, the operator for the most part achieves the QVE with these arbs

Recall that in Russian, obligatorily null 3PL arbs can be easily distinguished from mostly overt non-arbitrary 3rd-person plural pronouns. I utilised this fact in a corpus study of Russian to support the judgements in

When native speakers were presented with the latter examples in contexts that distinguished the two meanings, their judgements indicated lack of QVE.

As this pattern of readings indicates, 3PL arbs are never translated as indefinites in QVE contexts, but are strikingly similar to non-kind definite plurals.
This is supported by the lack of correlation between the readings of 3PL arbs and their sentential context (Condoravdi, 1989; Alonso-Ovalle, 2002; Cabredo-Hofherr, 2003). That is because the arb in the episodic sentence (31) can be interpreted as quasi-universal.

(31) Russian

Pervoje maja otmetili bujnym ves’el’jem
‘First May.gen marked.3PL wild.instr merriment.instr.’

‘The first of May was marked by wild merriment.’

This means that an existential translation for these arbs is unavailable, contra Jaeggli (1986); Condoravdi (1989); Kim (1991), Chierchia (1995b, p. 547), Cabredo-Hofherr (2003).

2.2.2 The semantics of 3PL arbs

Alonso-Ovalle (2002) proposes that generic uses of 3PL arbs in Spanish are definite plurals. In contrast, existential uses are derived using a subject-deletion operation. Such existential sentences “denote existentially quantified events without making any claim whatsoever about their agents” (Alonso-Ovalle, 2002).

Alonso-Ovalle’s proposal raises many issues. For instance, given the subject-deletion operation, how can we explain the presence of 3rd person plural agreement in (31)?

Cabredo-Hofherr (2003) offers an alternative analysis, in which the presence of a locative adverbial or a designated-subject predicate licenses a definite denotation for universal and corporate 3PL arbs, independently of the aspectual properties of the clause (31).

The existential readings arise through existential closure taking scope over the subject. To achieve these readings, the arb is translated as a free variable. This predicts that a Q-adverb should create the QVE with 3PL arbs, contrary to fact. Cabredo-Hofherr (2003) adduces the lack of QVE with these arbs to the 3rd-person feature of the pronoun: “since speaker and hearer have to be excluded, the unselective binding... cannot give an interpretation corresponding to people in general.” Thus, a more specific group excluding the speaker and hearer has to be pre-defined.

I raise two objections to this argument. First, it is not clear why exclusion of conversational participants from the range of a variable should prevent QVE. Indefinites exclude many individuals — e.g., a kid excludes adults and non-humans. Even NPs explicitly excluding participants yield QVE, as in (32).

(32) a. If people other than us deal with the Mafia, they usually get killed.

Thus, exclusion of conversational participants does not preclude QVE.

Second, 3rd person pronouns have been argued to involve no special participant-exclusion feature (cf. Harley and Ritter (2002)), and to be, in fact, unmarked for person (e.g., Benveniste, 1971, p. 217). Thus, in Malamud (2012a) I suggest features that yield a denotation the humans for 3PL arbs. It does not require exclusion of the speaker and hearer — rather, as with other 3rd-person pronouns, this exclusion arises as an implicated inference (Heim, 1991).
(but see D’Alessandro and Alexiadou (2003) who argue that 3pl arbs include [-speaker/-hearer] features). In fact, 3pl arbs in both English and Russian can express generalisations that include the speaker, as I discuss in the introduction. Thus, an English-speaking American can utter They speak English in America.

In sum, I doubt that QVE with bare-variable arbs can be prevented by a 3rd-person feature.

I propose, following insights in Alonso-Ovalle (2002); Cabredo-Hofherr (2003), that 3pl arbs are strongly definite on all of their uses. This may appear to be at odds with non-anaphoric and existential uses of 3pl arbs. However, non-arbitrary definites can also have antecedentless uses [33] that introduce new referents, like indefinites.

(33) The girls who attended a concert yesterday had fun.

Additionally, non-arbitrary definites can have non-maximal (existential-like) readings [34] [35]

(34) a. I started driving away from the mansion, but had to turn back — I’d left the windows open. (from Krifka 1996)

On this account, the anaphoric use of 3pl arbs is parallel to non-arbitrary pronouns. For other uses, I follow Siewierska and Papastathi (2011) who link the factors affecting the readings to referent identification. Setting aside the “saying” use of 3pl arbs, they group universal, corporate, and vague uses, in which the content of the arb is identified through linguistic context. These readings differ only in “the degree of overt referent identification... with... the referents of the universal being the most clearly identified (by the overt locative), that of the corporate less clearly identified (by the predicate and nonsubject argument), and... virtually no clue to identification being provided in the case of the vague” (Siewierska and Papastathi 2011, p. 584). For the inferred and specific existential uses, the referents are identified through extra-linguistic context: the result state from which the event is inferred, or through contact between the referent and the speaker.

Without providing a formal account, I suggest that this proposal is best interpreted as applying to the definite denotation of the arbs [17]

The universal and corporate uses arise essentially as proposed by Cabredo-Hofherr (2003). In the universal arbs, the locative expression supplies the group referent. On the four other uses, the existence of agents responsible for the described event is inferred from the verb. This inference could arise via a relational [agent] feature that is provided by the verb to its non-derived subject. As Alonso-Ovalle (2002) and Cabredo-Hofherr (2003) point out, the syntactic relationship between agentive verbs and their subjects is different from the relationship between verbs and derived subjects. Thus, just as in these proposals, the derived subjects are simply not in the correct configuration to receive a content-identifying feature

---

16 For three different approaches to deriving such non-maximal interpretations, see Landman (1989, 1996), Brisson (1998); Malamud (2006, 2012b).
17 Cabredo-Hofherr (2003, 2006) uses the term ‘content identification’ in the technical sense of Rizzi (1986): valuation of underspecified features. While I will ultimately suggest that this feature-based approach might be right, I do not mean to commit to a specific theoretical framework here.
from the verb, limiting all uses except the locative-licensed universal one to agents. The corporate arbs are further specified through cultural designation of a particular group of culprits.

Syntactic limitations on feature-transmission may also underlie cross-linguistic differences in the availability of different readings. Note that the vague, inferred, and specific uses differ in their aspectual properties: the vague involves existential quantification over events, the inferred use involves a perfect, and the specific use contains specific time reference. These aspectual distinctions correspond to syntactic differences. Thus, either the structures themselves, or permissibility of feature transmission in these different structures could vary cross-linguistically, yielding the typological hierarchy in [12]:

universal/corporate > vague > inferred/specific

(Siewierska and Papastathi, 2011).

The ultimate readings derived by this proposal are very close to those in Alonso-Ovalle (2002) and Cabredo-Hofherr (2003). The locative-licensed readings are available in generic and episodic sentences, and can be maximal (quasi-universal) and non-maximal (quasi-existential). In the absence of a locative expression, the use of the 3pl arbs is restricted to non-derived subjects. I leave the important details of the feature-transmission mechanism, and the way it plays out in various languages, to future work.

2.3 German man is indefinite-like

2.3.1 Man is susceptible to QVE with Q-adverbs

Unlike 3pl arbs, man produces the quantificational variability effects with Q-adverbs [35a], just as overt indefinites do [compare [35b]].

(35) German

a. An diese Fakultät ist man gewöhnlich klug / stolz auf den Dekan.
‘On this department is man usually smart / proud of the dean’
‘In this department one is usually proud of the dean.’
(QVE available: Most people in this department)

English

b. A Penn student is usually smart/proud of the dean.
(QVE available: Most Penn students)

Even more strikingly, the variable introduced with man in the if-clause of a donkey conditional can be bound by Q-adverbs [18], a hallmark behaviour of indefinites [compare 36].

18First, note that the presence of Q-adverbs makes these sentences into “generic” uses of man, in contrast to uses termed “existential” or “arbitrary”, and those termed “inclusive”. The purpose of the examples in [35] is to establish that these arbs cannot be analysed as uniformly definite, contra, e.g. Kratzer (1997). While it is possible that different denotations underlie different uses of man, below I will propose a uniform non-definite approach.
German
Wenn man klug ist, ist man gewöhnlich stolz.
‘If MAN smart is, is MAN usually proud’

‘If a person is smart, he/she is usually proud.’
(QVE available: Most smart people are proud)

English
If a man is intelligent, he is usually proud.

This pattern shows that man behaves like indefinites or kinds, both in the scope of Q-adverbs and in the if-clauses of donkey sentences.

The QVE with man cannot arise from a kind denotation. German man is not compatible with kind-selecting predicates in sentences such as [39].

German

(39) ??/* Eines Tages wird man aussterben.
‘Some day will MAN go.extinct’

Intended: ‘Someday, we/humans will be extinct.’

Zifonun (2000) proposes that man is a generic definite, but not a kind, because it lacks the intensional component of kinds (cf. Chierchia (1998), among others). If we take away the intensional component of a definite kind term, we are left with a non-kind definite denoting the maximal group of persons in the given situation. However, the only way in which this sort of denotation could exhibit quantificational variability with Q-adverbs would be to vary the situation itself, as in [20d], repeated below.

(20d) In the admissions process, we interview prospective students by inviting them into the room one at a time. If the kid is tall, he is usually smart.

Note that in the consequents of donkey conditionals in 18 man corresponds to a donkey-pronoun, like he in 36. I briefly discuss this exception to the indefinite-like behaviour of these arbs below; for a more in-depth discussion, see Malamud (2012a).

Hoekstra (2010) argues that man in Germanic languages, and impersonal pronouns in general, are kinds, citing the Frisian equivalent of [i].

i. Wenn man sich nicht fortpflanzt, wird man bald aussterben.
‘If MAN REFL not reproduce, will MAN soon go.extinct’

Intended, but does not mean: ‘If we humans don’t reproduce, we will soon become extinct.’

However, in German, [i] is not a statement about the humankind. Rather, it means that if members of some specific, contextually known group of people (including or excluding the speaker or hearer) won’t reproduce, they’ll soon be extinct. Thus, this is basically a QVE sentence, generalising over people within a contextually restricted set. In a survey I conducted, two other conditionals with man and kind-selecting predicates were rated as unacceptable.
If *man* is, in fact, such a definite containing a situation variable, what is the nature of this quantifiable situation? I leave the feasibility of the generic-definite analysis, and its ultimate details, to future research.

In sum, this means that in QVE sentences, *man*, like indefinites or properly set-up situational definites, introduces a new variable that Q-adverbs are able to quantify over.

2.3.2 The semantics of German *man*

The evidence laid out so far suggests that German *man* is not a strong definite (contra Kratzer 1997), and that it is not kind-denoting (contra Hoekstra 2010).

Moreover, two kinds of evidence suggest that *man* cannot be analysed as a regular quantificational indefinite.

First, in sentences with scope-taking operators, such as frequency adverbs and negation [40], the arb is scopally tied to the main eventuality expressed in its clause, and does not interact with other scope-taking items (Zifonun 2000; Cabredo-Hofherr 2010). Thus, [40a] cannot be interpreted to mean that some specific person repeatedly told me something, and [40b] cannot mean that some specific suspect failed to break in.

(40) German (Zifonun 2000, ex. (45b) and (52))

a. Man hat mir schon oft gesagt, dass...
   ‘MAN has me.dat already often told that’
   ‘I’ve often been told that...’

b. Man hat letzte Woche nicht bei uns eingebrochen.
   ‘MAN has last week not with us broken.in’
   ‘Our house wasn’t broken into last week.’

This is uncharacteristic of true (quantificational) indefinites, such as *jemand* “someone” [41].

(41) German (Zifonun 2000, ex. (46b))

Jemand hat mir schon oft gesagt, dass...
   ‘Someone has me.dat already often told that’
   ‘Someone often used to tell me that...’

This seems to be at odds with those generic (characterising) sentences where *man* has an existential interpretation [17]. I argue below that this contradiction is only apparent.

Second, unlike a regular quantificational indefinite, *man* can give rise to generic readings in an episodic sentence [16].

(16) Gestern feierte man in Baden-Württemberg “Heilige Drei Könige”.
   ‘Yesterday celebrated MAN in Baden-Württemberg “Holy Three Kings”’

---

21One possibility is that the situation variable associated with *man* is the situation associated with the verb phrase. However, such an analysis would not be able to derive the QVE, since it could not guarantee that we have only one person per situation. It is this one-to-one guarantee that enables weak definites to yield QVE in examples like [20].
‘People in Baden-Württemberg celebrated the holiday “Heilige Drei Könige” yesterday.’

Alonso-Ovalle (2002) uses sentences like [16] to argue against an indefinite interpretation for 3PL arbs. In a native-speaker survey I conducted, quantificational indefinites such as jemand were consistently judged to be unnatural in such contexts.

In Malamud (2012a), I analyse man as a bare variable, following prior work of Cabredo-Hofherr (2003, 2010), among others. The variable can be bound by a quantifier in the verbal domain (cf. Condoravdi (1989) on 3PL arbs). Here, I propose that this variable can also be associated with its own generic quantifier (cf. Moltmann (2006); Malamud (2012a) on English one, Malamud (2012a) on English you, German du). This proposal derives the behaviour of man on non-inclusive uses.

22 In Malamud (2012a) I build on this bare-variable core meaning of man to propose that inclusive man is, essentially, such a variable with an additional de se component. This allows speaker-inclusive readings in episodic sentences, and speaker-inclusive QVE readings in generic sentences.
In QVE sentences, and in plain generic clauses with quasi-universal *man*, the bare variable falls under the influence of the sentential quantifier.

In donkey conditionals [18], *man* behaves as an indefinite in the antecedent, and as a definite in the consequent clause, a problem originally discussed by Chierchia (1995b) for Italian *si*.

(18) German

Wenn man klug ist, ist man gewöhnlich stolz.

‘If MAN smart is, is MAN usually proud’

trans ‘If a person is smart, he/she is usually proud.’

In any theory of quantificational variability, an indefinite can be ‘bound’ only in the restriction of a Q-adverb, and existentially closed in other positions (de Swart, 1991; Chier-

---

In German, *es war* and *es gibt* existentials are claimed to involve a definiteness effect, whereby only indefinites may appear in the postverbal pivot position [i, is from Felser and Rupp 2001].

i. Es war ein Geist / *der Geist / *Elvis / *er im Schloss.
   ‘It was a ghost / *the ghost / *Elvis / *he in the castle’
   ‘There was a ghost/*the ghost/*Elvis/*he in the castle.’

Kratzer (1997) argues that strong unacceptability of *man* in these existentials is an instance of this effect.

However, Bayer (1986) points out that weak definites (generic, situationally variable, and kind definites), in general, are perfectly acceptable in existential constructions, arguing that the “definiteness effect” in German does not apply to definites. (I am grateful to an anonymous reviewer for pointing me towards this paper.) Moreover, many ‘stronger’ definites are perfect in existentials [ii, iii] (cf. Bayer 1986).

ii. *Why are the people in the village so sad?*

   A: Es ist gerade der Pfarrer gestorben. (from Bayer 1986)
   ‘it has just the clergyman died’
   ‘The clergyman just died.’

   iii. *Es gibt den Weihnachtsmann doch!*
   ‘it gives the Santa.Claus EMPH’
   ‘Santa Claus exists!’

Thus, definiteness would not rule out *man* from both types of existential sentences.

---

So, why is it ruled out? While a full explanation is beyond the scope of this paper, I suggest two avenues for a solution. First, Beaver et al. (2005) write: “pivot NPs are NPs that are not good candidates for functioning as subjects.” However, *man* is especially designed to be a good subject. For instance, it only has the nominative form (other cases being suppletive from *einer* ‘one’), and its discourse function is arguably to knock out the subject denotation from the salience / topichood computation (cf. Prince 2003, 2006 also §3 of this paper); thus it is a bad pivot. (Thank you to Gennaro Chierchia for suggesting this.)

Second, existential constructions force an existential interpretation for the pivot. In such contexts, *man* is restricted to be a non-derived subject (Zifonun, 2000) — that is, the subject of a transitive or unergative verb. However, as a pivot of an existential, it is doubtful that *man* is a non-derived subject, whether due to the properties of *war* (Grewendorf 1989), or to the properties of the pivot position. As I speculate in §3, this restriction may also originate in the pragmatic function of agent demotion associated with *man*. So, while my semantics for *man* does not preclude it from existentials, its pragmatic function does.
Otherwise, a sentence like (42a) might have an interpretation synonymous with (42b).

(42) a. If a guy is smart, a guy is rarely/usually proud.
    (no anaphora: the proud ones are not smart)

b. If a guy is smart, the guy is rarely/usually proud. (QVE: smart ones are proud)

Yet, man in the consequent clause is anaphoric to the previous occurrence of man. Since (18) involves variables co-varying under quantification, the two occurrences of the impersonals cannot be mediated by extra-linguistic reasoning (contra Koenig (1999); Koenig and Mauner (1999) for French impersonal on)24. Outside the realm of impersonals, only a definite can be a donkey pronoun, as is illustrated by the lack of anaphora in the consequent clause of (42a) which involves an indefinite a guy.

One way to resolve this challenge follows Chierchia (1998), who proposes that pronominal arbs are, essentially, bound by the same Q-adverb simultaneously in the antecedent and consequent clauses. The proposal invokes Principle C to rule out indefinites (but not indefinite pronouns) in the scope of Q-adverbs (42a)25.

In episodic sentences where man has a generic reading (16), the bare-variable analysis must assume an additional generic quantifier over the restrictor provided by the locative in Baden-Württemberg. This quantifier, while it is an additional stipulation in the analysis, may be justified by a phase in the historical development of the arb when it was restricted to generic contexts, like English one26.

In episodic sentences with existential man, the existential quantification in the verbal domain is responsible both for the episodic reading and the existential interpretation of the arb.

Finally, under this analysis, existential readings of man in generic contexts (17) are an illusion.

---

24 The use of arbs in consequent clauses of donkey conditionals looks like their so-called anaphoric uses, where they seem to co-refer with a use of the same arb in prior discourse. I suggest that extra-linguistic reasoning is at work here. The anaphoric reading is just a normal impersonal use, since whether it’s interpreted as co-referential with a previous arb is dependent on the coherence relation between the two sentences. The illusion of anaphoric use is, in fact, similar to examples such as [ii] where the non-anaphoric protesters can be interpreted to be the same group as people in [i] based on coherence. Of course, the bare plural in [ii] can also be referring to a new group, just as the arb can.

i. In New York, people marched to City Hall.  ii. There, protesters were arrested.

25 A problem with this is that definite descriptions are subject to Principle C [i], yet are ok in the consequents of donkey sentences (42b).

i. * A guy resembles a guy.

This problem is not specific to (Chierchia, 1998), but encountered by any extension of the E-type treatment of pronouns to other definites with bound-variable readings.

26 Egerland (2003) proposes that pronouns originating in a noun person follow the diachronic path in which they first become “generic” (only licensed in quantified sentences), and may go on to acquire “arbitrary” (existential) uses.
Da spricht man Spanisch.
‘There speaks MAN Spanish’
‘Spanish is spoken there.’

In these cases the clause contains an existential quantifier akin to sometimes or an existential modal akin to can. For instance, the sentence in [17] receives a modal interpretation (people in this store can speak Spanish), which gives rise to the existential reading for the bare-variable arb, just as it does in episodic-existential contexts. The existential quantification comes from the verbal domain, forcing the same scope for the main eventuality and man [40, 43].

(43) Scope inertness of man
    (after [Cabredo-Hofherr (2010)])

    Man hat mir schon wieder mein Rad geklaut.
    ‘MAN has me already again my bike stolen’

    ‘Someone has stolen my bike again.’ = ‘my bike was stolen again.’
    (most likely different thieves, i.e. again >> ∃)

Thus, man fails to interact with scope-taking elements (quantifiers or intensional operators), unlike a true indefinite.

The existential uses of man, like the corporate and existential uses of the Russian and English 3pl arbs, are subject to the non-derived subjects restriction (noted in [Jaeggli (1986)] for Spanish 3pl arbs, [Cinque (1988)] for French on and Italian si, and [Zifonun (2000)] for German man, among others). In the next section, I speculate on the historical origin of this restriction. Synchronically, an explanation should transcend various types of arbs. Under the bare-variable analysis of man, there is still a need to identify a restrictor set over which the variable ranges. This need can, as in the case of strongly definite 3pl arbs, be satisfied by an adverbial locative expression. For man, an additional mechanism is available with generic/sentential quantification, where the set serving as the restrictor for the sentential quantifier simultaneously provides the means of identifying the content of the arb. In the absence of these mechanisms, the denotation is simply the agents. I leave the details of deriving this denotation to future research, but suggest here, following [Egerland (2003); D’Alessandro and Alexiadou (2003); Cabredo-Hofherr (2004, 2010), among others, that this derivation might be tied to the feature specification of the arbs. Arbs of both types may contain a featural deficiency that, in the absence of the mechanisms mentioned above, requires valuation from the verb. Assuming, uncontroversially, that subject-related syntactic properties of agentive (unaccusative and transitive) verbs differ from the properties of non-

27 There is a fundamental distinction between the operators like “again” in [43] and Q-adverbs in QVE sentences: the former modify an existentially-quantified eventuality, creating an iterative interpretation, while the latter bind the eventuality directly. Thus the latter, but not the former, can yield the effect of quantifying over the arb.
agentive (derived-subject) predicates, it is easy to imagine that the special feature valuation can only be supplied by agentive verbs.

Restrictions on content identification for the restrictor set associated with man can be similar to those limiting the content of the definite \textit{3PL} arbs, especially if their source is a featural deficiency in the arbitrary pronouns.

### 2.4 Interim conclusions

I have presented evidence that different arbs pattern differently with respect to their susceptibility to quantification: \textit{3PL} arbs in English and Russian show QVE with quantifiers over parts, but not with Q-adverbs, while man in German yields QVE with Q-adverbs. This pattern suggests that the semantics of the definite-type arbs is quite different from the variable-type arbs, supporting the semantic dichotomy made in the typology proposed in [19a].

I suggest that \textit{3PL} arbs are definite plurals in both English and Russian, while German man is a bare, or generically-quantified variable. Thus, semantically arbs are not very different from non-arbitrary referential expressions. Prior researchers argued for a connection between the semantic and pragmatic properties of arbs \cite{Koenig1999, KoenigMauner1999, Egerland2003}, deriving both from the same underlying causes. In the next section, we turn to the pragmatic evidence for the typology proposed in §1.2 with a view to examine the nature of the connection. I will argue that pragmatically, arbs differ from non-arbitrary expressions.

### 3 Effects of arbs on discourse

Truth-conditionally, I have claimed that arbs are in some ways parallel to other definite and indefinite plural noun phrases. A natural question that arises is whether their discourse properties are those of non-arbitrary definites and indefinites. Note that dynamic frameworks like DMG \cite{Chierchia1995} or DRT \cite{Kamp1981, KampReyle1993} suggest a relationship between susceptibility to QVE and the potential of a referential NP to antecede discourse pronouns. To dwell briefly on this indirect relationship, definite NPs pick up old or

---

\footnote{What might be this feature deficiency? This is not an easy question. In discussing Italian \textit{si}, D’Alessandro and Alexiadou \cite{DAlessandroAlexiadou2003} write, “It is clear, as Chierchia (1995) suggests, that impersonal \textit{si} identifies a group of humans, but such a group needs further specification. This lack of specification corresponds to the lack of specification of the person feature on \textit{si}.”}

German man is, in fact, underspecified for person, since it is used for referents of 1\textsuperscript{st}, 2\textsuperscript{nd}, or 3\textsuperscript{rd} person. But a deficiency of 3PL arbs is not due to person: either 3\textsuperscript{rd} person more generally does not require person specification, or else it is already specified as [-speaker/-hearer]. Perhaps 3PL arbs have deficient descriptive/gender features. These arbs are specified to be human; however, Cabredo-Hoffêr \cite{Cabredo-Hoffêr2004} argues that this specification comes from the plural value of the number feature of these arbs, and not from descriptive features. Yet, the humanness of man cannot be attributed to plural marking, since man has none, and triggers singular agreement.

A theory that can simultaneously cut across the divide between the two types of arbs, and also respect the differences between their featural specification is an important task for future research.
uniquely identifiable referents; they correspond to an established discourse referent (cf. Heim 1982; Kamp 1981) and thus can serve as antecedents for subsequent anaphora. However, this potential to antecede discourse anaphora is not their primary function. In contrast, indefinites serve to exactly introduce new discourse referents for subsequent discourse anaphora — their susceptibility to quantification is one result of this, and their ability to antecede discourse anaphora is another.

As many authors had previously noted, the arbs that I claim pattern with semantic indefinites with respect to quantification are completely discourse-inert, and cannot antecede discourse anaphora. In this section, I turn to the effects that different types of arbs have on subsequent discourse pronouns. I report results that simultaneously corroborate the separation between the two types of arbs and support the pragmatic claims in 19b. I argue that this tension between 19a and 19b calls for a separation between semantic calculations such as the derivation of the QVE and pragmatic calculations concerning discourse pronouns.

Both types of arbs considered here have been compared to implicit agents in short passives, since they seem to function as ways of demoting the agent/subject denotation Myhill (1997); Prince (2003, 2006); Siewierska (2008a, 2010).

(44) The ship was sunk.

In this section, I turn to the effects of arbs and non-arbitrary NPs on subsequent discourse, and explore the degree to which the different arbs resemble or differ from implicit agents of short passives. In fact, definite-type and variable-type arbs affect discourse quite differently from non-arbitrary noun phrases, and from each other.

In contrast to German man, which cannot antecede a discourse pronoun, Koenig and Mauner (1999) show that implicit agents in English short verbal passives can, under certain circumstances, be subsequently followed by intersentential anaphora. This is true for 3PL arbs as well as for implicit agents, as the following constructed follow-ups to naturally occurring examples show. (Note that in Russian, conveniently, 3PL arbs must be null, while anaphoric pronouns must be overt. This the pronoun oni in 45a must be a non-arbitrary discourse pronoun. In contrast, the pronoun they in the second sentence of 45b may, in fact, be an arb.)

(45) Russian 3PL arbs

 a. Ivan uzhe privyk, chto ego rasskazu ne verjat ili ponimajut
   ‘Ivan already got used, that his story acc not believe.3pl or understand.3pl ego kak-to izvrashchennno. Oni ved’ ne vstrechalis’ s d’javolom. it.acc somehow perversely. They indeed not have.3pl met with devil.’
   ‘Ivan already got used to the fact that they don’t believe his story, or understand it somehow perversely. Unlike him, they haven’t met the devil.’

English passive

Koenig and Mauner (1999) argue that this co-specification between the implicit agent and subsequent pronoun is achieved by an inferential coercive process, rather than normal coreference. I disagree, and will argue against this proposal below.
b. The ship, made of iron, was cleaned and made environmentally safe before it was
sunk to the seabed. They wanted to create an artificial reef.

This referential possibility, however, is not evidence enough for declaring that the definite
arbs are identical to non-arbitrary NPs in their effect on subsequent discourse. Note that
effects on salience and topicality calculations can be gradient: various types of expressions
can form a salience scale. A semantic framework such as DRT is insufficient to model
such gradient effects, since it can only make one, categorical, distinction: an expression can
provide a potential antecedent for subsequent anaphora, or else it can fail to provide such
an antecedent.

To study salience and topicality effects, we need a framework that is, on the one hand,
compatible with dynamic or situation semantics, and thus will allow us to talk about the
results of the QVE-test presented in §2 and the discourse effects of arbs in a unified fashion.
On the other hand, the framework should allow us to detect fine-grained distinctions in the
salience and topicality of arbs and non-arbitrary pronominals. Both of these requirements
are satisfied by the framework of Centering theory.

The rest of this section presents the Centering theory (§3.1), and then describes corpus
studies of 3pl arbs in Russian and English (§3.2) as well as a study of German man (§3.3).
I then summarise the results and conclusions of these studies.

3.1 Centering theory

Centering theory (Joshi and Kuhn, 1979; Grosz et al., 1995; Walker et al., 1998) was proposed
as a model of local discourse coherence. The connection between salience and coherence
made in the theory was later the basis for pronoun-resolution algorithms (e.g., Brennan
et al., 1987). In the framework, the local discourse segment is broken up into utterances
(roughly equivalent to sentences). Several basic notions of the theory are defined as follows.

(46) **Cf-list** or the list of forward-looking centres is the set of discourse entities
evoked explicitly (and sometimes implicitly) in each utterance.

The entities on the Cf-list are ranked according to their salience, with language-specific
principles determining the ranking. In English and Russian, previous studies have deter-
dined the ranking as being roughly **subject** ≫ **object** ≫ **other** (Walker and Prince
1990; Malamud 2000, 2001), respectively). For German, Rambow (1993) proposes a
word-order dependent ranking; however, two corpus studies I conducted showed no evi-
dence against the **subject** ≫ **object** ≫ **other** ranking for German as well. Based
on the ranking, two entities can be selected as the most-likely future topic and the current
topic.

(47) ** Cp** or the preferred centre is the highest-ranked entity in the current utterance,
which is the predicted preferred topic for subsequent discourse, most likely to be
pronounalised in future utterances.

---

Thanks to Augustine Speyer, who was my collaborator for the first of these studies.
CB or the backward-looking centre is the most salient (highest-ranked) entity from preceding utterance realised in the current one. This is the centre of the hearer’s attention, the actual topic of an utterance, linking this utterance to preceding discourse.

The backward-looking centre is the item most likely to be realised as a pronoun in the current utterance. Based on the change and retention of backward-looking centres and preferred centres, transitions of different levels of coherence are defined between adjacent utterances.

For example, in [49a] below, the same backward-looking centre (John) is retained in utterance 2 and 3, and the Cp of utterance 3 is also the same entity. This is the smoothest type of transition, called Continue. Another transition type in which the Cb is retained is illustrated in [49b]. While the backward-looking centre is the same, the Cp is a different entity, which means that the transition between the utterances is Retain, the second-smoothest type.

Next, in [49c], while the Cb and Cp of the third utterance are the same, indicating a probability that the speaker will continue talking about Mary, there is a change of backward-looking centre between utterance 2 and utterance 3, making this transition a Smooth-Shift. The roughest type of transition, the Rough-Shift is illustrated in [49d], where the topic (i.e., the backward-looking centre) changes from utterance 2 to utterance 3, and the projected topic for subsequent discourse (i.e., the preferred centre) is also different from the Cb.

(49) 1. John went to school this morning. 2. He met Mary on his way.
   a. 3. He was in a good mood.
   b. 3. Mary often walked to school with him.
   c. 3. She carried a pink school-bag.
   d. 3. Bill had been annoying her.

As numerous Centering studies indicate, grammatical subjects have a special status in the calculation of topicality. In the languages considered here, they are the preferred centres in their utterances — they are most likely to be pronominalised in subsequent discourse. Thus, an effect of a passive or a sja-passive is to remove the implicit agent (logical subject) denotation from this prominence, thereby demoting its influence. Both types of arbs considered here also share a special connection to the subject position. Specifically, 3pl arbs are predominantly used in subject position; man is also predominantly a subject. Thus, they are positioned to exert a strong influence on the topicality and salience calculation of subsequent discourse.

To examine the influence of arbs on discourse, I conducted corpus studies of their Centering behaviour, and the patterns of pronominalisation and pro-drop in discourse segments subsequent to a use of an arb. I turn to these now.
3.2 3PL arbs in discourse

Utilising 3PL arbs has the effect of decreasing the topicality or salience of the arbs’ plural reference. This becomes apparent in sentences containing a competing antecedent for plural anaphora. Even nonsensical overt antecedents provide strong competition for arbs. [50]

(50) Russian

Zdes’ pekut horoshije pirogi. Oni starajutsja ugodit’ klientam.
‘Here bake.3PL good.acc pies.acc. They try.3PL to.please clients.dat’

‘They bake good pies here. They try to please the clients.’

A very salient (and for 6 out of 10 native speakers asked, the preferred) reading of [50] is one where the pies are trying to please the clients. Of course, overt and sensible antecedents are almost unbeatable competition. [51]

(51) Russian

Zdes’ vybirajut horoshih artistov. Oni starajutsja ugodit’ klientam.
‘Here choose.3PL good.acc performers.acc. They try.3PL to.please clients.dat’

‘They choose good performers here. They try to please the clients.’

The only reading in [51] is the one where the performers, and not their employers, try to please the clients. Thus, the pronoun (which refers to the most central, the highest-ranked salient entity from the preceding utterance) is preferentially (and in [51], categorically) resolved to something other than the arbitrary subject. Given that, in general, subjects in Russian are typically the likeliest antecedents for future anaphora (and likeliest topics for subsequent discourse), we can conclude that definite plural arbs are different from the typical Russian subject. In Centering terms, we say that these arbs succeed in placing a discourse entity on the ranked list, thus making future anaphora possible, but that the entities are ranked extremely low on the list, making future anaphora very much dispreferred.

To confirm this judgement data, I have conducted a corpus study of the status of 3PL arbs in Russian literary texts (twentieth and twenty-first century fiction and non-fiction) collected from the online library of Maxim Moshkow (www.lib.ru) and the Russian National Corpus (see RNC). In addition, I conducted a smaller corpus study of 3PL arbs in a genre-matched English corpus.

A Centering analysis of the Russian corpus data showed that 3PL arbs are almost never actually followed with intersentential anaphora (7 cases out of 1003 items in the Russian corpus).

Moreover, under the assumption that the phonologically null Russian arbs are backward-looking centres, like the personal null pronouns in pro-drop languages, we predict that the discourse segments containing the arbs have an unusually high number of incoherent (Rough-Shift) transitions. A chi-squared test showed that there is a less than 1% chance that the different distribution of transitions in the texts with and without arbs is due to chance. In contrast, an analysis that only counted non-arbitrary pronouns as especially salient predicted
a much higher degree of coherence. This prediction coincided with the intuitive perception of the discourse by two native speaker readers. This corpus study suggests that while agents of sentences involving arbs can provide antecedents for future anaphora (at least when this is the only way to make sense of a discourse), they are not preferred topics nor highly salient items.

English results are much harder to interpret, because a 3\textsuperscript{rd} person plural pronoun following an arb is ambiguous between an impersonal and a personal interpretation. I therefore counted the proportion of cases when a 3\textsuperscript{pl} arb was followed by another 3\textsuperscript{rd} person plural pronoun in the next utterance: this occurred in 32.5\% of cases. I then produced three Centering analyses of the segments containing the arbs: one that considered the arbs to be central and maximally-salient in their utterances (representing both the Cb and the Cp), and two analyses that placed the arbs on the bottom of the Cf list. These two analyses that considered the arbs to have low salience differed in their treatment of the subsequent-mention 3\textsuperscript{rd} person plural pronouns: one analysis treated them as arbs, and thus ranked them at the bottom of the Cf list (that is, low-salience for the purposes of Centering), and the other analysis considered them to be personal discourse pronouns, and thus potential centres of attention.

In segments containing no such subsequent mention, the transitions ignoring the arbs were, on average, considerably smoother than those that considered the arbs to be salient. There were no cases where the arbs-as-centres analysis produced a better transition than the low-salience, analysis. The situation in segments containing subsequent, ambiguous, 3\textsuperscript{rd} person plural pronouns, is a bit more complicated. In the vast majority of cases, the subsequent pronouns could be interpreted as co-referential with the arbs. Thus, considering both the arbs and the following pronouns to be the backward-looking centers resulted in a smoother transition for the utterance following the arb. However, since the arb itself does not refer to any entity from prior discourse, an analysis that considers the arb to be the centre of its utterance always produces a smooth-shift transition for the utterance containing the arb. In contrast, an analysis that ignored the arb often yielded the more coherent retain and continue transitions for the utterance containing the arb. As a result of one analysis having an advantage for transitions following the arb, and the other analysis having an advantage for transitions preceding the arb, on average, neither analysis predicted a more coherent discourse for these segments.

Note that utterances where the arb received no subsequent mention predominated, constituting over two-thirds of all arbs. I take this fact to indicate that arbs do not serve to introduce new, salient entities into discourse. Moreover, in the vast majority of cases where there was a subsequent mention, it did not persist beyond one utterance. Together with the overall smoother transitions calculated using an analysis that considered arbs to have low salience, I take this lack of arbs’ persistence in subsequent discourse to indicate a similar role to that of the Russian 3\textsuperscript{pl} arbs: the arbs place their denotations at the bottom of the salience-ranked list.

Since I claim the 3\textsuperscript{pl} arbs to be semantically similar to definite plurals, this raises the possibility that perhaps their level of salience is similar to that of lexical definite descrip-
tions. Such a study would be difficult to conduct in Russian, which does not have articles. Therefore, to explore this possibility, I conducted a smaller corpus study for overt definite descriptions in an English corpus (comprised of several Wall Street Journal and other short-form and long-form journalism, as well as fiction writing). Previous Centering studies of English (Brennan et al., 1987; Prince, 1992; Walker and Prince, 1996, among others) showed that subjects are preferred topics in this language. However, when only subjects that are plural definite descriptions with the are considered, we find that they are pronominalised quite rarely in subsequent discourse. Overall, definite plural subjects received subsequent mention in 27% of cases. This observation is confirmed in the studies conducted by Michaelis and Francis (2004) on a corpus of spoken English, where lexical (as opposed to pronominal) definite noun phrases are pronominalised in subsequent discourse a third less than lexical indefinites. While 21% of the definite descriptions in my corpus were the centre of attention of their utterance, most were not. However, the Centering analysis suggests that in 88% of cases, the definite plurals were the highest-ranked elements of their utterances, and thus the preferred topics for future discourse. The discourse segments containing these definite plural subjects were fairly smooth in terms of the distribution of transitions: with 24% smooth-shifts, 46% retains, and the rest continues, with only a single rough-shift. The transitions from prior discourse to the target utterance are, for the most part, retains (56%). An analysis which reclassified all definite plural subjects as highest-ranked in their utterances, that is, as Cps, resulted in 64% retain, 24% continue, and 12% smooth-shift transitions from prior discourse to the target utterances. This is much smoother than transitions to the arb utterances, if the 3pl arbs are classified as Cps. Thus, in stark contrast to the segments with 3pl arbs, an analysis that considered non-arbitrary definite plurals to be highly salient did not produce an unusually high proportion of shifts.

While a much larger corpus is necessary to achieve statistical significance for these results, the two corpus studies strongly suggest several important conclusions about the role of arbs in discourse. First, in English, the lexical non-arbitrary plural definite descriptions in subject position are normally not centres (Cbs) of their utterances. Nevertheless they are still strongly salient, just like indefinite and pronominal non-arbitrary subjects.

While English 3pl arbs are about as likely to receive subsequent mention as non-arbitrary definite plural subjects, the latter are highly salient elements of their utterances (Cps), and thus strongly influence the salience and topicality calculations in discourse. In contrast, both English and Russian 3pl arbs are not highest-ranked in their utterances. Thus, the denotations of arbs have a very weak effect on the salience calculation, and other discourse referents outrank the arbitrary subjects.

In sum, I have shown that 3pl arbs in English and Russian are reduced in salience, more so than non-arbitrary definites. However, they still exert some (weak) influence on the salience/topicality calculation.

An alternative explanation for the possibility of anaphora to the arbs, as in [50], might be that 3pl arbs fail to place their referents on the ranked list; the anaphora would then be made possible through accommodation (Lewis, 1979). This is exactly what Koenig and

31 Thanks to an anonymous reviewer for pointing out this project (and its online location) to me.
Mauner (1999) propose for implicit agents of short passives. However, this proposal cannot work — as I will observe in the next subsection for German man, and as Koenig (1999) observes for French on, there are arbs that indeed fail to place anything on the list. For such arbs, accommodation is not an option, as evidenced by impossibility of discourse anaphora. Yet, I see no way to prevent the general inferential reasoning posited by Koenig and Mauner (1999) from applying with these arbs as well.

Thus, we need to distinguish between the low-salience items, such as 3PL arbs in English and Russian, and items that do not participate in the salience calculations at all, such as the German man, to which I now turn.

3.3 Man in discourse

In contrast to 3PL arbs, it has been noted by many researchers that man cannot provide antecedents for discourse pronouns of any kind (Kratzer 1997; Zifonun 2000; Egerland 2003, among others). Prince (2003) reports on a Centering study of a Yiddish corpus containing me(n), the Yiddish correlate of German man; this is expanded in Prince (2006) to include the French dedicated impersonal pronoun on. The patterns of pro-drop and pronominalisation force a conclusion that me(n) and on fail to place an entity onto the ranked list; that is, the agent denotation in sentences with me(n) and on is never considered to be a potential topic for subsequent discourse, or a potential antecedent for intersentential anaphora.

I conducted a replication of Prince’s studies for a corpus of German, consisting of both written texts (fiction and non-fiction), as well as spoken data (transcripts of radio programs and European Parliament sessions). From this corpus, I extracted 100 segments where utterances with man were followed by an utterance containing a discourse pronoun or, in five segments, a short anaphoric definite description (altogether 25 segments from each genre). The reference of the pronouns and definite descriptions was resolved using judgements of five native speakers for each segment. I then subjected these segments to Centering analysis, an example of which is given in (52) below.

(52) “Die westliche Kirche vom 10. bis zum frühen 12. Jahrhundert” (The western Church from 10th through early 12th century) by Gerd Tellenbach
   a. In den meisten christlichen Ländern war der Papst eine erhabene, fast
      ‘in the most Christian countries was the Pope a sublime almost
      legendäre Gestalt in weiter Ferne.
      legendary figure in wide distance’
      ‘In most Christian countries the Pope was at a distance, an almost legendary
      figure.’
      Cb: the Pope
   b. Von seiner realen Existenz wußte man im allgemeinen wenig...
      ‘Of his real existence knew MAN in common little’
      ‘One generally knew little of his actual existence...’
      Cb: the Pope
   Transition ignoring man: Continue; Transition including man: Retain

35
Among the Christians in Muslim Sicily, the Pope is said to have been almost forgotten.

There were zero instances of discourse anaphora referring to man. In each case, the entity that would have been ranked second after the agent was the most likely to be dropped or pronominalised in subsequent discourse. Discourse segments that included a string of sentences with man were equally as likely to have co-referential interpretation for the arbs as they were to have non-co-referential instances of man. In addition, analysis that ignored man had overall much smoother transitions than analysis that included man. This evidence from transitions, pronominalisation and pro-drop showed that an entity other than the agent denotation in sentences with man was the preferred centre.

This provides a striking contrast with overt indefinites, which are exactly the providers of new discourse entities to the ranked list of potential topics.

Given these Centering results, I conclude that the chief discourse function of man is to remove the subject/agent denotation from the topic-structure of its sentence (the list of forward-looking centres), allowing the next-highest-ranked entity to become the topic in the Centering sense.

What is the source for this pragmatic impotence? Egerland (2003) compares human impersonal pronouns Scandinavian man, French on, and Italian si with 2nd-person impersonals and Icelandic maður. His theory relies on the fact that human impersonal pronouns he examines share the following properties:

- they can trigger plural or singular agreement,
- they may have a generic or an arbitrary (existential) interpretation, and
- the arbitrary (existential) use is precluded in episodic (‘specific time reference’) contexts with ergative or passive verbs, and, finally,
- these pronouns cannot be syntactic objects.

He goes on to explain these properties through a featural deficiency of the impersonals in question: they “radically lack inherent lexical content with regard to the categories of person and number (and presumably also gender),” except for a [+human] feature restricting the denotation to humans. This deficiency then also serves to explain their inability to serve as antecedents for discourse pronouns.

In contrast to the arbs Egerland examines, German man provides evidence against a theory that links together these properties. On the one hand, man allows both generic and arbitrary readings and in episodic sentences cannot be a subject of passives or ergatives.
man would demote the subject denotation, removing it from the Cf list. In 53a, the subject is a theme/patient, not agent. In contrast to the unaccusative example in 53b, where the subject is the agent, the existential reading is not available for the arb in 53a, suggesting that man is specially designated for agent demotion (cf. Kallulli 2007).

(53) German (from Cabredo-Hofherr 2004, p10, ex[29] 32

a. #Gestern wurde man verprügelt.
   ‘Yesterday was MAN beaten.up’
   Does not mean: ‘Yesterday someone was beaten up.’

b. #Gestern kam man um 10 Uhr an.
   ‘Yesterday came MAN at 10 hour to’
   Does not mean: ‘Yesterday someone arrived at 10 o’clock.’

Note that in generic and quantificational contexts, man can be a non-agent (as in 52 above). However, if we interpret man as falling under the influence of the sentential quantifier in such sentences, the quantifier will prevent the arb from introducing a top-level discourse referent, which thus will not need to be demoted separately. I speculate that the agent-demotion function of man is historically responsible for the restriction of man to be a non-derived subject in existential cases. Siewierska (2011) argues that such restrictions are characteristic of higher degrees of grammaticalisation, and existential uses are more grammaticalised than the generic ones 33.

On the other hand, German man does not have the flexibility with respect to gender and number that Egerland observes for Scandinavian man, French on, and Italian si. With respect to verbal agreement 34, nominal concord, and possessive pronouns, German man is always singular and masculine (Zifonun 2000). Despite its ability to antecede reciprocals, man cannot support plural nominal/adjectival concord 54.

(54) German

In diesem Institut ist man gewöhnlich ein schlauer Mensch / *schlaue Leute.
   ‘In this institute is MAN usually a smart person / *smart people’
   Intended: ‘In this institute, a person is usually smart.’

The upshot is that the discourse-effects of German man cannot be explained through mere feature deficiency. Here, I note these effects, and leave their explanation (and the

32 However, an anonymous reviewer disagrees with these judgements. This is in line with the wide range of dialectal variation amongst speakers of German as reported in Malamud (2012a).

33 Note that both a pragmatic framework such as Centering Theory I utilise in this section, and a semantic one are needed to describe this possible historical development: the quantification in generic uses prevents top-level referents from arising at the semantic level, while any remaining top-level referents are then demoted in the pragmatic calculation. This would explain why generic uses are immune from this restriction.

34 Egerland reports data for Swedish, which does not have agreement on finite verbs. His examples, therefore, all involve participles, which may involve a different mechanism underlying agreement. Arguably, this different mechanism would be the same for participial agreement and for nominal concord (cf. ?)
underlying featural composition of arbs) to a different investigation (cf. Malamud (2012a) for a discussion of the feature composition of *man*).

### 3.4 Pragmatic conclusions: agent demotion in arbs

The corpus studies reported in this section, and specifically the rareness of subsequent pronominalisation of definite-type arbs, and the absence of pronominalisation or pro-drop for variable-type ones, support the analysis in which the definite arbs are placed on the bottom of the salience hierarchy (the Cf list), making subsequent discourse anaphora possible yet extremely dispreferred. At the same time, the variable arbs would not be placed on the list of potential antecedents at all, making subsequent discourse anaphora entirely impossible, as suggested in Prince (2003, 2006) for Yiddish *me(n)* and French *on*.

This study of the discourse effects of arbs provides further evidence for the separation between the definite-type arbs, which place their referent on the bottom of the Cf list, and the variable-type arbs, which fail to place anything on this list. Thus, both pragmatic and semantic evidence point to the separation that I argue for in the proposed typology.

The Centering studies allow us to measure the agent-demotion function that has long been attributed to both types of arbs (Myhill, 1997; Siewierska, 2011). As Prince (2003, 2006) notes, the arbs provide a way of precluding the outsize effect that subjects have on salience calculations, short of syntactically removing them from the subject position. She writes, regarding the discourse-inert impersonal pronouns in her study:

> “Notice that what we have here is a kind of asyntactic passive, a functional equivalent of the passive, a lexical means for ‘demoting’ the subject and making it, in Relational Grammar terms, a chômeur, i.e. an inert argument that cannot be manipulated syntactically, at the same time as the object (indirect or direct) is ‘promoted’ to the discourse equivalent of subject in that it is construed as the center of attention and what the next utterance will be about.” (Prince, 2006)

However, the existence of the pragmatic distinction between the two types of arbs raises the question, what is the discourse effect of passivisation? Koenig and Mauner (1999) suggest that implicit agents in passives are discourse-inert, like *man*. In the next section, I will argue against this, pointing out that syntactic passives are closer to 3PL arbs than to discourse-inert *man*.

### 4 Implicit agents in the typology of arbs

The typology I proposed is based on only three items in three different languages. This naturally raises the question whether it can be consistently expanded to other arbs. In this section, I present preliminary evidence that it is indeed so expandable, specifically addressing implicit agents of short verbal passives in Russian and English and implicit agents of
Russian *sja*-passives. These items seem to pattern with definite-type arbs.\(^{35}\)

(55)  
(a) Russian short passive

Vchera byl potoplen vrazheskij korabl’.

‘Yesterday was sunk enemy.nom ship.nom’

‘Yesterday, an enemy ship was sunk.’

(b) English short passive

Yesterday in America, Independence Day was celebrated with joy.

(56) Russian *sja*-passive

Novyj god prazdnovalsja dolgo i radostno v etot raz.

‘New.nom year.nom celebrated.SJA long and joyfully in this time’

‘New Year was celebrated long and joyfully this time around.’

Unlike the 3PL arbs and *man*, in passives and *sja*-passives, the patients and not the implicit agents serve as subjects. While the form of Russian short verbal passives is a mirror of its English counterpart, involving a copula and a passive participle of the (formerly) main verb, verbal morphology in Russian *sja*-passives is not passive, but reflexive. In all these passive constructions, the agent is not overtly expressed, and the patient occupies the subject position and agrees with the verb. The agent, however, is present in the semantic representation. For instance, it can control the PRO in purpose clauses.\(^{57}\)

(57) Russian

Komnaty zdes’ provetrivajutsja, chtoby uluchshit’ kachestvo vozduha.

‘Rooms.nom here air.SJA to improve quality.acc of.air’

‘Rooms here are being aired to improve the air quality.’

In formal semantics, implicit agents have usually been assigned an existential, indefinite denotation.\(^{1978}\). However, note that implicit agents in episodic sentences can be understood as almost-universal among relevant individuals, as in \(^{58}\), an interpretation unexpected for an indefinite.

(58) Russian (from RNC)

(a) Prazdnovalsja kakoj-to jubilej... v restoran’e “Osen”

‘Celebrated.SJA some jubilee... in restaurant “Osen”’

---

\(^{35}\)Elsewhere, I presented evidence that Italian *si*-impersonals pattern with variable-type arbs, both in their susceptibility to QVE and in their effects on discourse.\(^{[Malamud 2006]}\).

\(^{i}\) Italian *si* impersonal (from Chierchia 1995b, p. 107)

In Italia, si beve molto vino

‘In Italy *si* drinks much wine’

‘They drink lots of wine in Italy.’
b. Some jubilee was being celebrated in the restaurant “Osen”.

In fact, contrary to the common assumptions that implicit agents in short verbal passives are indefinites \cite{Dowty1978}, placing them in QVE contexts shows that they pattern rather with overt definite plurals \cite{59}:

\begin{itemize}
\item \textbf{English}
\begin{itemize}
\item a. In Spain, Michael Jackson is for the most part admired.
\hspace{1cm} (QVE available: Most Spaniards = QVE on Spaniards)
\item b. In Spain, Michael Jackson is usually admired.
\hspace{1cm} (QVE not available: the only reading is ‘Now he’s admired, now he’s not.’)
\end{itemize}
\item \textbf{Russian}
\begin{itemize}
\item a. V Rossii, “Pravda” bol’shej chast’ju preziraetsja.
\hspace{1cm} ‘In Russia, “Pravda” is for the most part despised.’
\hspace{1cm} (QVE available: Most Russians = QVE on Russians)
\item b. V Rossii, “Pravda” obychno preziraetsja.
\hspace{1cm} ‘In Russia, “Pravda” is usually despised.’
\hspace{1cm} (QVE not available: the only reading is ‘Now it’s despised, now it’s not.’)
\end{itemize}
\end{itemize}

Similarly, the implicit agents in Russian sja-passives pattern with strong definites in QVE contexts \cite{60}:

\begin{itemize}
\item \textbf{Russian}
\begin{itemize}
\item a. Dver’ byla obychno otkryta.
\hspace{1cm} ‘The/a door was usually open(ed)’
\hspace{1cm} ‘The/a door was usually open(ed)’
\hspace{1cm} Does not mean: Most people opened the door.
\end{itemize}
\end{itemize}

It is important to note that this variability is truly dependent on the presence of the implicit agent. This is illustrated by the absence of any such effect in adjectival passives \cite{61}, where neither the quantificational adverb, nor \textit{for the most part} have the agent denotation to quantify over.

\begin{itemize}
\item \textbf{English}
\begin{itemize}
\item a. The door was usually/for the most part (un)opened.
\hspace{1cm} Does not mean: Most people did(nt) open the door.
\end{itemize}
\item \textbf{Russian}
\begin{itemize}
\item b. Dver’ byla obychno otkryta.
\hspace{1cm} ‘Door was usually open(ed)’
\hspace{1cm} ‘The/a door was usually open(ed)’
\hspace{1cm} Does not mean: Most people opened the door.
\end{itemize}
\end{itemize}

Thus, the QVE test seems to group implicit agents in English and Russian short passives and sja-passives with 3PL arbs.

Is this grouping consistent with the Centering test — that is, do these arbs pattern together with respect to their effect on salience and topicality calculations?
In contrast to *man* and similar to 3PL arbs, implicit agents permit subsequent pronominal reference in certain contexts. This is most clearly seen in Russian, where the overt pronoun cannot be interpreted impersonally, and thus must represent discourse anaphora. The following made-up continuations of naturally-occurring examples are acceptable (see also the example in [45b]).

(62) Russian

a. Tol’ko odno voshozhdenie bylo sdelano s severa. Oni razrabotali novyj marshrut cherez lednik.
   ‘Only one ascent was made from the north. They developed a new route across the glacier.’

b. Na trave mojetsja krasnaja loshad’. Oni skrebut jej golovu
   ‘On the grass a red horse is being washed. They’re scrubbing its head...’

However, not all passives and *sja*-passives can be felicitously followed by a pronominal reference to the implicit agent. In an acceptability rating task with 7 naïve native speakers, the average acceptability of such continuations to sentences with no competing antecedents for the subsequent pronoun seems to be significantly lower than the corresponding statistic for 3PL arbs. While a larger study involving a corpus investigation is necessary for a definitive conclusion, I can nevertheless suggest that matching features of the 3PL arb and the subsequent pronoun facilitate coreference between them. In contrast, it is not clear that implicit agents have any features at all, except perhaps those restricting them to certain thematic roles, and thus cannot provide a feature match for subsequent pronouns. Thus, discourse anaphora to implicit agents involves both co-reference and accommodation of the requirements imposed by the pronominal features.

A Centering analysis of 50 segments from the Russian National Corpus demonstrates that implicit agents are outranked by subjects, as well as other entities mentioned in the passives and *sja*-passive. A made-up continuation for the RNC example in [63] demonstrates that the plural pronoun *oni* can only refer to the posters, not to the agents of the passive.

(63) Russian

Ves’ gorod byl obklejen afishami. Oni reklamirovali novyj spektakl’.
‘Entire city was pasted over with posters. They advertised a new show.’

‘The entire city was pasted over with posters. They advertised a new show.’

This evidence supports the proposed typology, since the implicit agents pattern with 3PL arbs with respect to Centering: their referents are placed on the bottom of the list of potential reference, making anaphora possible but dispreferred.
What are the consequences of this classification, grouping implicit agents with definite arbs? Alonso-Ovalle’s proposal for 3pl arbs makes their existential uses similar to passives, in that they are achieved by deleting the subject denotation altogether (in the context of existential quantification over the event). My tests make generic uses of 3pl arbs similar to passives, as well, in that QVE is generated with “parts” and not with Q-adverbs. Perhaps implicit agents of passives have, in fact, definite denotations similar to the agents, but unspecified for number and humanness. In this case, the existential readings of implicit agents are simply non-maximal readings of these definites.

An observation that seems to go against this proposal is that there is an important difference between 3pl arbs and implicit agents: while existential readings for the latter seem to be available in all languages and contexts, for the former existential readings may be, at times, unavailable. However, the features of the two arbs are quite different: implicit agents have no features other than possibly those causing restriction to certain thematic roles, while 3pl arbs are human, definite, plural, perhaps exclusive of conversational participants. Thus, if restrictions on content identification are responsible for restricting availability of existential readings, it is plausible that those restrictions would differ for the different feature combinations.

The grouping of 3pl arbs and implicit agents of passives into the same typological category also makes sense diachronically. In many languages, 3pl arbs, in fact, grammaticalise into passives (Siewierska, 2010). Their shared properties set the stage for such diachronic change.

While a more thorough study is necessary to investigate the consequences of this data for the semantics and pragmatics of implicit agents, their placement in the present typology suggests a re-visioning of the traditional indefinite approach to these constructions.

In sum, while both types of arbs demote the agent denotation, it is the definite 3pl arbs, and not the variable man that best deserve the term “asynatactic passive” (Prince, 2006). The implicit agents of passives, while they are even less likely to antecedate pronouns than 3pl arbs, function identically to these arbs with respect to Centering. This supports the subtle differences between the discourse effects of definite-type arbs and variable-type arbs, as well as the correlation between the semantic and pragmatic parts of my proposed typology. The existence of pragmatic distinctions between the two types of arbs and non-arbitrary NPs, together with the semantic differences between them has important consequences for the architecture of the semantics-pragmatics interface. I turn to the consideration of these consequences now.

5 The typology and the semantics-pragmatics interface

In a series of studies Koenig (1999); Koenig and Mauner (1999); Koenig et al. (2002) and others investigated the effects of passives and impersonals on subsequent discourse. The studies tied these pragmatic effect to the semantic representations of implicit agents and dedicated impersonal pronouns, working in the Discourse Representation Theory (DRT) (Kamp and Reyle, 1993). Both types of arbs, according to Koenig and collaborators, satisfy
the argument slot of the predicate, but do not participate in the referential structure. That is, they claim that these items fail to introduce a new discourse referent or to pick up a discourse referent introduced earlier. In this section, I argue that a purely semantic framework, like the one employed by [Koenig (1999)], is not rich enough to capture the discourse-relevant distinctions between non-arbitrary NPs, variable-type arbs like *man*, definite-type arbs like *3pl* impersonals, and implicit agents of passives.

As [Prince (2006)] points out, one of the main tasks of a semantic framework like DRT is to account for the entailments of sentences. In particular, we must account for the QVE facts: definite-type arbs show QVE with *for the most part* but not Q-adverbs, while variable-type arbs yield QVE with Q-adverbs. In DRT, the distinction can be captured by a built-in difference between the discourse referents introduced by different arbs. Definite-type arbs like *3pl* ones and, according to the results in the last section, implicit agents of passives, behave just like non-arbitrary strong definites with respect to QVE. In DRT, these expressions must introduce a top-level discourse referent that is pegged to a pre-existing group of people, and thus is unavailable for quantification. In contrast, variable-type arbs, such as German *man* and, according to the results reported in [Malamud (2006)], Italian *si* must introduce a variable, either a bare variable ranging over people, or a situation variable as a part of the weak definite. It is this variable, a new discourse referent in DRT terms, that enables the Q-adverb to create the effect of quantifying over the arb.

Thus, if, as Koenig and colleagues claim, neither impersonal pronouns nor implicit agents introduce a discourse referent of any kind, there is no way to capture the semantic part of the typology. On the other hand, once the discourse referent is introduced, and its properties match the QVE behaviour of the arb, there is no mechanism that could prevent its participation in discourse anaphora.

Conversely, an analysis that casts the pragmatic effects of arbs as inability to introduce discourse referents fails to capture the pragmatic part of the typology. Specifically, the evidence I present requires at least five distinctions combining semantic and pragmatic properties of expressions (cf. [Prince (2003, 2006)] and discussion therein). First, non-arbitrary indefinite subjects satisfy the argument slots of predicates, entail the existence of their referents, introduce a variable susceptible to QVE, and provide highly salient potential antecedents for future discourse. Second, non-arbitrary definites share all these properties, but their referents are not, without additional context, susceptible to QVE. Third, definite-type arbs (including implicit agents) satisfy the argument slots of predicates, generate existence entailments, are not susceptible to QVE, and provide very low-salience potential antecedents for future discourse. Fourth, variable-type arbs satisfy the argument slots of predicates, generate existence entailments, introduce variables for QVE, and do not provide any potential antecedents for discourse anaphora. Fifth and finally, quantificational expressions, such as *nobody* satisfy the argument slots of predicates, but arguably generate neither generate existence entailments nor QVE, and do not provide potential antecedents for discourse pronouns.

A semantic framework such as DRT or DMG is not sufficiently rich to support these five distinctions. The proposal in [Koenig (1999)] makes two distinctions. First, definite and indefinite non-arbitrary expressions introduce familiar and new discourse referents, respectively.
Second, arbs, like quantificational expressions, saturate the predicates but fail to introduce discourse referents. If neither variable-type nor definite-type arbs introduce a discourse referent, then subsequent reference to implicit agents and 3PL arbs can only be explained as accommodation. In this case, variable-type arbs such as German *man* must somehow be able to prevent accommodation, since these arbs cannot be followed by discourse pronouns. While the mechanisms enabling accommodation in general are poorly understood, nevertheless restrictions on these mechanisms do not depend on specific lexical items which have identical contributions to the semantics — rather, there are general inferential processes that can facilitate or discourage accommodation. This means that we have no way of simultaneously permitting accommodation for 3PL arbs and ruling it out for *man*.

The solution is one I have adopted in presenting the typology of arbs: we must separate semantic calculations (in frameworks such as DRT, DMG, or any other) from pragmatic ones (in frameworks like Centering). Only then can we model semantic behaviour, such as definiteness, quantificational variability, or existence entailments, and simultaneously track fine-grained salience distinctions on the pragmatic level. In the DRT terms, both types of arbs introduce a discourse referent into the semantic representation. Those introduced by definite-type arbs are pegged to a specific referent unsuitable for QVE, and marked as low-salience for the pragmatic calculations. At the same time, the referents introduced by variable-type arbs contain a new variable available for QVE, and are marked as unavailable for pragmatic calculations.

The correlation between the semantic and pragmatic parts of my typology predicts that the two types of calculations (semantic and pragmatic) are not completely independent, since in that case we would expect arbs that are definite-type with respect to QVE while being unavailable for anaphora; as well as arbs that are variable-type on the semantic level while permitting (though discouraging) anaphora. Broadening this inquiry to more arbs in more languages should tell us whether these are possible.

To conclude, I presented evidence from corpus studies and native speaker surveys for a meaning-based typology of items with arbitrary interpretation (arbs). The typology distinguishes two types of arbs based on their semantic and pragmatic properties. One consequence of the proposal is an impossibility of a single, uniform semantics for the two types of arbs. Another is a possibility that implicit agents of passives may be, at least in some contexts, definite, contrary to common assumptions about their meaning.

Finally, I present an argument that semantic and pragmatic calculations must be separate, since a semantic framework that captures the similarities between arbs and non-arbitrary indefinites cannot also account for the subtle pragmatic differences between them. This argument extends and develops the line of research presented in Prince (2006), addressing the question whether a single model of referring expressions can explain both their truth-conditional and discourse properties. She examines two *man*-type impersonal pronouns in Yiddish and French, and concludes that their discourse-inertness cannot be derived in a framework that remains faithful to their semantics. The present investigation broadens the inquiry to include additional evidence that variable-type arbs introduce variables
that influence the truth-conditions in a way that differs from the referents of definite-type arbs. The semantic contributions of arbs are shown to be in some ways similar to those of non-arbitrary definite and indefinite NPs. At the same time, the contributions of the two types of arbs to the pragmatic salience/topicality calculations differ from each other, and from non-arbitrary NPs. The resulting salience scale calls for a more fine-grained model of reference at the level of pragmatics than at the level of truth-conditions, and thus supports Prince’s conclusion that semantic and pragmatic models must be separate.

Acknowledgements

I am grateful to Elsi Kaiser for inviting me in 2005 to submit to the volume on impersonals for which this paper was originally intended, and to Anna Siewierska for encouraging me in the fall of 2010 to publish this research in a journal. Additional collection of corpus and judgement data for the revisions of this paper were partially funded by the Computer Science department of Brandeis University. I am deeply grateful to my informants, too numerous to name, to my family and especially Robert for support throughout, to anonymous reviewers of both the volume and the journal versions of this paper for helpful and thoughtful feedback, and to Anna Slavina for editing support. This paper is dedicated to the memory of Ellen Prince who taught me more than I can say.

References


