

What ‘You’ and ‘I’ Mean to Each Other:

Person Indexicals, Self-Ascription, and Theory of Mind

Stephen Wechsler

University of Texas

Department of Linguistics

1 University Station B5100

Austin, TX 78712-0198

wechsler@mail.utexas.edu

In *Language* 86 (2): 332–365 .

What 'You' and 'I' Mean to Each Other:

Person Indexicals, Self-Ascription, and Theory of Mind

Abstract

This paper offers a *de se theory* of person indexicals, wherein first and second person indexical pronouns indicate *reference de se* (also called *self-ascription*). Long observed for first person pronouns (Castañeda 1977; Perry 1979; David Kaplan 1977, *inter alia*), self-ascription is extended here to second person as well. The person feature of a pronoun specifies the speech act roles that must be played by the self-ascribers: the speakers (uttering a first person pronoun), the addressees (interpreting a second person pronoun), or both (for first person inclusive). Other agents who are not among the designated self-ascribers for a given pronoun interpret the pronoun indirectly by inferring the self-ascriber's interpretation, a process requiring *theory of mind*, i.e. the cognitive ability to impute mental states to others (Premack and Woodruff, 1978). This *de se* theory is supported by convergent evidence from multiple domains: (i) It explains a typological universal: first and second person plurals always allow associative semantics ('speaker(s) plus others'; 'addressee(s) plus others') rather than requiring regular plural semantics ('speakers only'; 'addressees only') (Greenberg 1988, Noyer 1997, Cysouw 2003, Bobaljik 2008). (ii) It belongs to a family of approaches that solve the problem of the essential indexical (Perry 1979). (iii) It correctly predicts observed patterns of indexical pronoun production and comprehension by two populations lacking a fully developed theory of mind: typically developing children in the stage before theory of mind has developed; and children with autism. (iv) It correctly predicts the interpretation of second person pronouns in utterances with multiple addressees.*

1. Introduction

Extensive typological studies of pronominal person/number paradigms have revealed a striking universal generalization (Moravcsik 1978, 356; Greenberg 1988, 14; Noyer 1992, 31; Cysouw 2003; Siewierska 2004, 82-83). First and second person plural pronouns always have *associative* semantics rather than *regular plural* semantics. That is, second person plural pronouns are never grammatically restricted to refer to addressees alone, but rather their reference sets can include others as well; and first person plural pronouns are never restricted to multiple speakers, but can include others as well. Past explanations for this *Associative Plural Generalization*, as it will be called here, involve stipulating a small set of universal person features with a special associative, rather than regular plural, semantic interpretation (Silverstein 1976; Noyer 1992; Bobaljik 2008).

The alternative explanation for the Associative Plural Generalization presented here does not depend upon such a stipulation. Instead it is rooted in the philosophical and semantic accounts of person indexicality as *self-ascription*, also called *reference de se* or simply *self-reference*. It is proposed that both first and second person pronouns encode self-ascription, where the person feature indicates who is to self-ascribe the property of being the referent of the pronoun, the speaker or the addressee, respectively. With a plural pronoun, the speaker or addressee self-ascribes membership in the reference set of the pronoun. First person (exclusive) pronouns designate the speakers as the self-ascribers, second person pronouns designate the addressees as self-ascribers, and for first person plural inclusive pronouns, both speakers and addressees are self-ascribers.

Other interpreters of a given pronoun, such as addressees hearing a first person pronoun, speakers uttering a second person pronoun, or eavesdroppers, interpret these pronouns indirectly by inferring the self-ascriber's interpretation. That inferential process involves imputing mental states to other people and making predictions based on those imputed mental states, an exercise of the cognitive ability known as *Theory of Mind* (Perner, 1991; Premack and Woodruff, 1978; Schneider et al., 2005). The present proposal receives important support from studies of pronoun use by two groups lacking a fully developed theory of mind: children at an age before theory of mind has fully developed (roughly before the age of 3.5 to 4 years); and children with autism, who are hypothesized to have a theory of mind deficit. Both groups have special problems with first and second person pronouns, including a tendency to reverse them.

The paper is structured as follows. Section 2 presents the Associative Plural Generalization, then shows that it is an absolute universal (§2.2), that it is semantically strange (§2.3), and that previous accounts are not explanatory (§2.4). In Section 3 the De Se Theory of indexicals is proposed and defended. By way of background, an issue for indexicals known as the *problem of the essential indexical* (Perry 1979) is presented first (§3.1). The proposed solution builds on Asher (1986) and Crimmins (1992) (§3.2). A semantics is sketched for singular (§4.1-2) and plural (§4.3) first and second person pronouns, supported with preliminary linguistic evidence (§4.4). Section 5 shows that the Associative Plural Generalization is an automatic consequence of the De Se Theory, constituting what is perhaps the strongest evidence for that theory. The De Se Theory is then further supported with evidence from the acquisition of pronouns by typically developing children (§6.2) and pronoun use by children with autism (§6.3).

2. Universal properties of personal pronoun systems

2.1 The Associative Plural Generalization

Certain strange properties of personal pronoun systems, which have been revealed through extensive cross-linguistic studies to be absolute universals, receive an immediate explanation from the *de se* theory of indexicals proposed below. In this section we present those universal properties and review the previous attempts to explain them.

First and second person singular pronouns refer to the speaker and addressee, respectively. But it has long been noted that first and second person *plural* pronouns do not necessarily refer to a collection of speakers or addressees, but rather to a collection of people that *includes* the speaker or addressee. Hence these so-called plurals have associative rather than regular plural semantics. Let us consider this in detail.

English first person plural pronouns (*we, us, our, ourselves*), for example, can be used to refer to any group of individuals that *includes* the speaker or speakers. That group may also include the addressee(s) and/or others who are neither speakers nor addressees.

(1) Possible interpretations of English first person singular and plural pronouns

<u>interpretation</u>	<u>example</u>
a. 'speakers'	sg: <i>I need a drink.</i>
	pl: <i>We are the champions! (unison)</i>
b. 'speakers + others'	<i>We want you to come to dinner.</i>
c. 'speakers + addressees'	<i>Shall we go?</i>
d. 'speakers + addressees + others'	<i>Can't we all get along?</i>

It is rare for multiple speakers to speak in unison (see below for discussion), but we include that possibility for completeness. The plural term 'speakers' in 1 should be understood as 'speaker or speakers'; *mutatis mutandis* for 'addressees' and 'others' in 2 and 3 below.

Similarly, a second person plural pronoun can be used for reference to any group that includes all the addressees, a group that may, but need not, also include others who are neither speakers nor addressees.ⁱ

(2) Possible interpretations of English second person forms.

a. 'addressees'	sg: <i>You should behave yourself.</i>
	pl: <i>You should behave yourselves.</i>
b. 'addressees + others'	<i>How do you guys handle promotions over in Philosophy?</i>

Many languages have distinct first person plural pronouns for including the addressee (first person *inclusive*) or excluding the addressee (first person *exclusive*). Such languages have one form comprising interpretations 1a,b, and another for interpretations 1c,d. Examples include Indonesian (exclusive *kami*, inclusive *kita*) and the Nilo-Saharan language So (exclusive *isia*, inclusive *inia*).

(3) First person plural exclusive (3a,b) versus inclusive (3c,d)

<u>interpretation</u>	<u>example (Indonesian)</u>
a. 'multiple speakers'	<i>kami</i>
b. 'speakers + others'	<i>kami</i>
c. 'speakers + addressees'	<i>kita</i>
d. 'speakers + addressees + others'	<i>kita</i>

The possible interpretations of personal pronouns are summarized in Table I below. Using the traditional numerical symbols, 1 for speaker, 2 for addressee, and 3 for other, we can, following Bobaljik (2008), define the seven logically possible 'meta-persons' in the column labeled 'Possible' in Table I (this table is based on Bobaljik 2008:205). Noyer (1992:160) and Bobaljik (2008:205) observe that 'certain distinctions are never morphologized'. The maximal attested system is the four-way contrast shown in the column labeled 'Attested'. Languages like English make even fewer distinctions, lumping together inclusive and exclusive.

Despite considerable cross-linguistic variation in personal pronoun systems (Cysouw 2003), certain logically possible distinctions are not found. As Noyer

(1992:161-2) puts it, no person category specifically *excludes* 3; and no category allowing speaker or addressee *requires* 3. In other words, as long as the value of the number feature is appropriate, including an ‘other’ (someone not participating the speech act) in the denotation of a pronoun is always possible, and never required— apart from third person pronouns, which exclude speaker and addressee entirely.

INSERT Table I ABOUT HERE

Bobaljik (2008:206) provides these two different characterizations of the universals, based on the three missing meta-persons:

(4) Person universals

a. As restrictions on contrasts

U1. No language distinguishes [1+1] from [1+3].

U2. No language distinguishes [2+2] from [2+3].

U3. No language distinguishes among [1+1+2], [1+2+2] and [1+2+3].

b. As restrictions on forms

U1. No language has a special morpheme for (true) [1 PL] .

U2. No language has a special morpheme for (true) [2 PL] .

U3. No language has a special morpheme for ‘comprehensive’ person [1+2+3].

Taking U1 first, no language has a ‘choral *we*’ pronoun specialized for referring exclusively to multiple speakers ([1+1], that is, ‘true [1 PL]’). The lack of such a pronoun is perhaps not so surprising, since it could only be used for mass speaking in

unison. But a number of potential applications have been noted, including chanting at sports events (*We are the champions!*), ritual mass speaking as in a church service, children at play, and so on (Cysouw 2003:73-4, citing Mühlhäusler and Harré, 1990: 201-2). Nevertheless, Cysouw (2003:74) observes that ‘As far as I know, however, there is no language in the world that distinguishes a separate morpheme for mass speaking.’

Universal U2, which applies to second person, is much more surprising. A pronoun specialized for referring just to the addressees and no others (2+2) would seemingly be very useful. Nevertheless: ‘However great the semantic plausibility, the category 2+2 is not found grammaticalized in the languages of the world.’ (Cysouw 2003:75). For example, the southern American English second person plural *y’all* can refer either to a plurality of addressees (2+2) or to a group consisting of one (or more) addressee plus other non-addressees (2+3). The same is true for all other languages, as far as we know.

Having described the Associative Plural Generalization, we will now show the following: (i) that the Associative Plural Generalization is supported by very strong cross-linguistic evidence, lending support to the claim that it is an absolute universal (Section 2.2); (ii) that it is a strange fact, given the way that the semantics of plurals works otherwise (Section 2.3); and (iii) that previous explanations offered for it amount to directly stipulating it within Universal Grammar (Section 2.4). Then in Section 4 we show that Associative Plural Generalization follows automatically from the present *De Se Theory* of indexical pronouns.

2.2 The Associative Plural Generalization is an absolute universal.

The quotes from Cysouw in the previous section attest to his confidence that the Associative Plural Generalization is a true universal: ‘no language in the world’ violates U1, and the form prohibited by U2 ‘is not found grammaticalized in the languages of the world.’ Cysouw’s conclusion should carry some weight. It is based on a study of over 309 languages from diverse genetic and geographic groups (Cysouw 2003).

Moreover, Cysouw’s study is only one among several large-scale typological studies that reached the same conclusion. Bobaljik’s (2008) metastudy of those studies concludes that the three the Associative Plural Generalization universals are ‘absolute universals rather than strong trends’ (Bobaljik 2008, 209). Bobaljik (2008, 207) observes that ‘the empirical basis is extremely well-documented’, citing ‘sample sizes on the order of 500 languages’ from studies over the past half century (Forchheimer 1953; Sokolovskaja 1980; Cysouw 2003). In another influential book, Siewierska concurs that ‘The two interpretations of the second-person plural [i.e. 2+2 and 2+3 — S.W.] appear not to be formerly (*sic*) distinguished in languages (see Moravcsik 1978, 356; Greenberg 1988, 14; Cysouw 2003, 71). Nor is the rare 1+1 reading of the first-person plural.’ (Siewierska 2004, 82-3).

Some exceptions to the Associative Plural Generalization have occasionally been claimed (Bobaljik 2008:208 cites Comrie 1980; and Plank 1985). But those challenges have been surveyed and ultimately dismissed in three recent works (Simon 2005; Cysouw 2003; Bobaljik 2008). A few key issues are briefly summarized here, closely following Bobaljik 2008, but the reader is referred to those three works for more detailed discussion.

First, the universals hold of monomorphemic person markers. Some languages have semantically transparent compound pronouns, such as the English-based creole Tok Pisin inclusive form *yumi*, from ‘you’ plus ‘me’. It refers to the speaker and addressee but no others. The Associative Plural Generalization claim is that forms with such meanings are never morphemicized.

Second, a number feature, in combination with a person feature, can effectively draw the distinctions prohibited by the universals, but these are not distinctions within the person system itself. For example, obviously the singular number feature, together with first or second person, restricts reference to only the speaker or hearer, which has the effect of excluding reference to ‘other’. Subtler cases arise in languages with more non-singular number distinctions. Dual number plus inclusive person has a similar effect of including the (sole) speaker and (sole) addressee, but excluding ‘others’.

An illustration of the complicated effects of the number feature is found in Ilocano, which appears at first blush to distinguish pronouns for [1+2] (*ta*) from [1+2+3] (*tayo*) (Cysouw 2003, Bobaljik 2008). Crucially, *tayo* can refer to the speaker plus multiple addressees, hence it is not really a [1+2+3] form, but rather a [1+2+X] form, where X can be anyone (Bobaljik 2008, 218-219). So it is not an exception to the Associative Plural Generalization. The two forms are both inclusive, and are distinguished by number, not person. The form *ta* was traditionally claimed to encode dual number, but the language lacks independent motivation for the dual number value. A different analysis treats the Ilocano number system as distinguishing *minimal* from *augmented* number: ‘minimal’ is used for the minimal number of individuals necessary to satisfy the meaning of the pronoun, apart from number; ‘augmented’ indicates more

than that minimal number. Within this system *ta* is minimal and *tayo* augmented. Siewierska (2004, 84-85) sketches a similar minimal versus augmented analysis of the pronoun system of Uradhi.

More generally, to assess the Associative Plural Generalization against descriptions of particular languages, it is necessary to determine whether or not a pronoun's reference set can potentially include (non-speaker, non-addressee) 'others'. Some descriptions are vague on this point, but when the issue is clarified through more careful study, the apparent counter-examples vanish. See Cysouw 2003, Simon 2005, and Bobaljik 2008 for detailed discussion of particular cases.

2.3 The Associative Plural Generalization is semantically strange

The upshot of the Associative Plural Generalization is that for first and second person forms, pluralization yields *associative* meanings rather than the usual plural semantics. First person singular refers to the speaker, but first person plural refers, not necessarily to a multiplicity of speakers, but rather 'the speaker plus associates'. Second person singular refers to the addressee, but second person plural refers, not necessarily to a multiplicity of addressees, but rather 'the addressee plus associates'. Nunberg (1993, 8) finds that this is a 'curious meaning for a plural noun phrase to have', suggesting that 'We would be surprised, to put it mildly, if we ran across a hitherto undescribed dialect of English that had a plural form *doggen* that meant "a group of animals that includes a dog", as in "Doggen are most tranquil when they are not also catten."'

More recent work suggests that Nunberg overstated his case. Associative morphemes for proper and even common nouns are found in many languages, e.g. the

Japanese associative marker *-tachi*, as in *Tanaka-tachi*, ‘Tanaka and his family or friends or associates’; or the Hungarian associative marker *-ék*, as in *Péter-ék* ‘Peter and his family or friends or associates’ (Moravcsik 2003, 469). Moravcsik (2003, 469) notes that ‘while associative plurals are of restricted distribution in any one language, they are fairly widespread across languages.’ She points out some semantic similarities between plural pronouns and associative NPs, and concludes that they are essentially the same, differing only in the type of nominal, pronoun versus noun.

Given the existence of both *regular plural nouns*, with their very broad distribution, and *associative nouns*, with ‘restricted distribution in any one language’ (Moravcsik 2003, 469), the question for us is why the plural features of first and second person pronouns are always semantically interpreted like the latter, highly restricted associative nouns, and never like the former, relatively unrestricted regular plurals. First and second person pronouns are *never*, in any language, specified exclusively for normal plural semantics. We don’t even find languages in which normal plural and associative plural pronouns coexist. This situation contrasts with noun forms, as in Hungarian distinction between associative plural *János-ék* ‘John and associates’ and regular plural *János-ok* ‘Johns (more than one person called John)’ (examples from Corbett 2000, 102). First and second person pronouns always have associative semantics when they are plural. That fact is a mystery that this paper addresses.

While that mystery is heretofore unsolved (except by stipulating that it is part of the innately endowed Universal Grammar; see Section 2.4), there have been some attempts at the more modest goal of rendering it less strange. Let us consider to what extent that more modest goal has been achieved.

Corbett (2000, 83ff; 2005, 10) attempts to minimize the difference between nouns and 1st/2nd person pronouns, when it comes to associative readings of plurals. Following Smith-Stark (1974), Corbett posits an animacy hierarchy of nominal and pronominal categories according to their likelihood to have plural forms in a given language, where items on the left are most likely and items on the right are least likely to allow plurals.

(5) speaker > addressee > 3rd person > kin > human > animate > inanimate

Different languages make the split at different points in the hierarchy. For example, in many North American languages, only nouns referring to human beings, and pronouns, have plural forms (Mithun 1988, 212, cited in Corbett 2000, 58); in other languages, such as Maori, it is mainly kin terms such as *matua* ‘parent’ and *teina* ‘younger sibling’ that show number marking, while other human-denoting nouns do not (Corbett 2000, 60-61).

More directly relevant here is the further claim by Corbett that this animacy hierarchy governs the relative likelihood of different interpretations of the plural number feature. Associative interpretations are most likely to be found at the far left of the hierarchy, regular plurals most likely in the middle, and the option of recategorization of mass versus count terms at the right end. Thus since ‘speaker’ and ‘addressee’ are at the far left, they tend strongly towards associative readings of the plural, but such readings are also possible with other nouns, according to Corbett (2000, 84; 2005, 10):

We see, therefore, that the fact that associative readings are found most readily with personal pronouns forms part of a more general pattern, and does not

indicate that personal pronouns are quite different in respect of number. If we were to treat them differently, then we would have to give up the typological regularities based on the Animacy Hierarchy. (Corbett 2005, 10)

That is, Corbett suggests that some plural nouns are like Nunberg's hypothetical, surprising *doggen* that means "a group of animals that includes a dog". Being at the left end of the hierarchy, the first and second person pronouns would then favor associative readings, but would remain on a continuum with nouns.

But Corbett does not give many examples of associative readings of plural common or proper nouns. In fact the only example I could find in either of the two works cited is from English, and appears in the context of the following comment: 'Associative readings with plural nouns are relatively rare in English but are found with plural anaphoric reference, as in *Aunt Doris rang while you were out. – Oh, how are they?*' (Corbett 2005, 10, fn. 14) But this example does not contain a plural noun, nor does it convincingly illustrate associative semantics of the plural pronoun *they*. An associative plural should have some descriptive content applying to one member of a group. For example, *we* has the content 'speaker', *you.PL* has the content 'addressee', Japanese *Tanaka-tachi* ('Tanaka and associates') has the content 'named Tanaka', and so on. But the pronoun *they* lacks such descriptive content. In Corbett's example the pronoun *they* can refer to a group that includes Aunt Doris because that group is contextually salient (Heim and Kratzer 1998, 240). (The associative semantics are not essential to this type of anaphora resolution; cp. the interpretation of *she* in *I met a couple yesterday that you might know. — Oh, is she Norwegian?*) As far as I can tell, as applied

to associative readings of plurals, at least in English, the putative animacy hierarchy is marked by a very sharp divide, with first and second person pronouns displaying associative readings of plurals on one side, and everything else showing regular plural readings on the other.

As noted above, some languages like Japanese and Hungarian have specialized associative morphemes. These apply most often to proper nouns, and in some languages common nouns accept them as well (Moravcsik 2003). Corbett (2000, 2005), following Corbett and Mithun (1996), argues that such associative morphemes are not part of the number system, observing that the associative variable cuts across different non-singular number values in languages distinguishing dual from plural number. Corbett (2005) further argues that plurality in first and second person *is* part of the number system. Moravcsik (2003), on the other hand, argues that so-called first and second person plurals are literally associative forms. Regardless of how that issue is resolved, the same question remains: Why do first and second person pronouns lack ‘plural’ forms with the regular plural semantics characteristic of other nominals? (If second person singular means ‘addressee’, why is there no second person plural meaning ‘addressees’?) For Corbett the question is why those first and second person plurals lack regular plural semantics; for Moravcsik the question is why first and second person pronouns have associative forms but no regular plural forms.

While Corbett attempts to assimilate pronouns to nouns, Schlenker (2003a, 6; 2003b, 54) attempts to assimilate person to gender. He notes that ‘in French the masculine plural pronoun *ils* need not denote a group that only includes males; rather, the requirement is simply that at least one member of the group be male.’ (Schlenker 2003a,

6) This seems similar to first or second person plural: ‘at least one member is the speaker/addressee’. Schlenker (2003a, 6) proposes ‘a semantics that is formally similar’ for gender and person.

But there are two different ways to state the gender rule for French (Corbett 1991, 279-280):

- (6) Two types of gender rule (Corbett 1991, 279-280)
- A. 1. If at least one member of the group is masculine, the masculine form is used;
 - 2. otherwise the feminine is used.
 - B. 1. If all members of the group are feminine, the feminine form is used;
 - 2. otherwise the masculine is used.

Schlenker assumes a type A statement for gender, which is necessary in order to assimilate person to gender. But for gender resolution in most languages, including French, the evidence favors type B; while in other languages, a type A rule seems to be indicated.

French masculine gender is used on a predicate adjective when the subject lacks any gender feature at all, as when it is a clause or infinitival phrase. That shows that masculine, not feminine, is the default gender applying when no rule pre-empts that default, as in rule B above. Also masculine is normally used when the speaker doesn’t know the gender. Languages with more than two genders provide even clearer evidence for distinguishing between rules of type A and B. Slovene and Serbian/Croatian have three-gender systems, with masculine (M), feminine (F), and neuter (N). When all

elements are feminine (F+F), then *feminine plural* is used. *Masculine plural* is used for all other combinations of genders, including M+F, M+N, N+F, M+M, and even N+N. The Slovene and Serbian/Croatian rules are thus identical to 6B, but with a different effect due to the third gender. This system cannot be stated with a Type A rule.

A very common resolution system involves a general distributivity rule stating that if all members belong to gender γ , then use the plural of γ ; with a special default or resolution gender for gender mixtures (Corbett 1991, Ch. 9; Dalrymple and Ronald M. Kaplan 2000; Wechsler 2008). Icelandic works that way, with neuter as the resolution gender; hence M+F, for example, gives neuter plural. Bantu languages work that way as well (Givón 1970). If all conjuncts have the same gender (Noun Class), then the verb appears in the plural of that Noun Class; Noun Class mixtures are resolved to a specific gender for humans (the 1/2 form), and a different one for non-humans (e.g. the 7/8 form in Luganda). All such languages require rule statements of Type B; none allows a Type A rule.

On the other hand, some other languages seem to employ a type A rule.ⁱⁱ Polish has a total of four genders, with feminine, neuter, and two subgenres of masculine, namely masculine personal and non-personal. However, there are only two plural target forms, masculine personal and non-[masculine personal]. The Polish rule states that if at least one conjunct is masculine personal, then the masculine personal form is used; otherwise the non-masculine personal is used.

Summarizing, gender systems tend to use type B resolution rules, where the gender feature effectively distributes over the members of the set and a default gender is

designated for resolving heterogeneous sets. But some languages use type A rules, where the inclusion of a certain gender among the elements is sufficient to trigger its use.

The *person* features, in contrast, *always* employ type A rules, to wit this system (based on Corbett 1991, 262):ⁱⁱⁱ

- (7) a. If the group includes a speaker, then first person is used;
- b. if the group includes an addressee, then second person is used;
- c. otherwise third person is used.

Or, in a system with an inclusive/exclusive distinction:

- (8) a. If the group includes a speaker and addressee, then inclusive is used;
- b. if the group includes a speaker, then exclusive is used;
- if the group includes an addressee, then second person is used;
- c. otherwise third person is used.

What would a type B system for person look like? It would violate the Associative Plural Generalization. A system with a rule like ‘if all members of the group are addressees, use the [2PL] form’ is exactly what the Associative Plural Generalization prohibits; *mutatis mutandis* for speakers and [1PL]. To formulate a type B analysis of English person would require negatively defined categories like ‘non-speaker’ and ‘non-addressee’, which lack independent motivation from the language.

In conclusion, despite Schlenker's attempt to assimilate the semantics of person features to that of gender features, the two categories behave very differently. Plural person features have associative interpretations in all languages, while plural gender features only occasionally do. Similarly, the associative semantics found with person features is attested with nouns, as expressed with associative morphemes, comitatives, and perhaps even some uses of ordinary plural morphology (Moravcsik 2003). So associative semantics is an available option in language. But the question remains as to why this option becomes the only one available for first and second person pronouns. The Associative Plural Generalization remains semantically strange, and cries out for an explanation.

2.4 The Associative Plural Generalization and Universal Grammar

Following Silverstein's (1976) pioneering work, Noyer (1992), Bobaljik (2008), and others analyze the attested types of person system with cross-classifying features. The presentation in this section closely follows that of Bobaljik 2008. Two universal boolean features are shown in the column of Table II labeled 'binary features': $[\pm\text{spk}]$ indicates whether or not the speaker is included in the referent of the pronoun, and $[\pm\text{hr}]$ (for 'hearer') indicates whether or not the addressee is included.

INSERT TABLE II ABOUT HERE

If Universal Grammar makes available only these two features, and allows only those interpretations of the features, then the Associative Plural Generalization is expected.

Bobaljik further argues that alternative functional and conceptual explanations are unable to provide satisfactory accounts of the Associative Plural Generalization. His conclusion, that the feature system responsible for the typological universals is specific to human language, forms the basis for a seemingly powerful argument for substantive linguistic universals. The elimination of alternative explanations would lend support to Noyer's (1992, 168) assumption that the feature system is 'part of Universal Grammar.' This type of account of the Associative Plural Generalization will be referred to as a *UG solution*.

The UG solution just sketched encodes the Associative Plural Generalization formally. But it does not really explain the generalization. It does not explain why the *person* feature is unique among semantic features in requiring a special associative semantics, where [+spk] means 'speaker plus associates', and [+hr] means 'addressee plus associates'. As discussed in the previous section, gender features as well as other semantic features typically operate precisely by prohibiting 'associates': *the dogs* refers to a set containing only dogs and no non-dog associates; feminine, as a marked gender, is typically used for females and not for male associates. What makes the category of person so special?

It is also significant that the Associative Plural Generalization appears to be an absolute universal. Other putative substantive universals from UG are either (i) very general properties of language structure that can be plausibly attributed to basic properties of human cognition or interface conditions, such as recursion in syntactic structure; or (ii) more specific substantive constraints on grammar, but constraints that admit of exceptions among languages. The latter type is the basis for theories of parameterization. Such 'universals' function as default settings that can be overridden in

acquisition by positive evidence of exceptionality. But the Associative Plural Generalization does not appear to have such exceptions. The lack of exceptions, together with the unique behavior of person features as opposed to other types of feature, suggests a deeper explanation that is rooted in the nature of the person feature itself, that is, in the semantics of indexicality. Such an explanation is provided below.

To get a sense of what it would mean for an explanation of the Person Marking Generalization to be ‘deeper’ than the UG solution above, suppose counterfactually that some language is found to have a ‘true second person plural’ that refers to all and only the addressees. Such a language would obviously falsify the UG solution, but only in a shallow sense. Under the UG solution such a language could be accommodated by positing an extra feature such as [–other] to indicate absence of ‘other’, or a special language-specific interpretation of the [+hr] feature. That featural adjustment would have no consequences for other aspects of the grammar. In particular, this hypothetical language would not require a different semantics of indexical pronouns. But under the account proposed below, it would. The Associative Plural Generalization follows as a logical consequence of the *de se* theory of indexical pronouns proposed in the next section.

3. The *de se* theory of indexical pronouns

3.1 The problem of the essential indexical

An important insight from the philosophy of language and linguistic semantics from the second half of the twentieth century is that indexical expressions— forms whose reference depends directly upon the utterance context, such as first and second person pronouns— are *essential* in the sense that they cannot be semantically reduced to non-

indexical descriptions (Perry 1979; Kripke 1979; Castañeda 1977; David Kaplan 1977; Richard 1983; Lewis 1979; Geach 1972). Perry (1979: 3) illustrated this *problem of the essential indexical* with his tale of the spilled sugar:^{iv}

I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally it dawned on me. I was the shopper I was trying to catch.

In the course of this event, the beliefs of the speaker, John Perry, undergo a change ('Finally it dawned on me'), as shown by his bending down to fix the torn sugar sack in his shopping cart. Perry (1979) considers how he can linguistically report that new belief with sentences of the form *X is making a mess*, such as the following.

- (9) a. I am making a mess.
 b. The shopper with the torn sack is making a mess.
 c. John Perry is making a mess.
 d. He [*pointing to a reflection of himself in the mirror*] is making a mess.

These sentences have the same propositional content, but only 9a unambiguously expresses Perry's new belief. Apart from the first person pronoun, no other designation of Perry unambiguously captures that new belief in a way that would explain why he

straightened the sugar sack in his cart. Clearly substituting 9b will not explain his behavior, since he knew from the outset that ‘The shopper with the torn sack is making a mess’. Replacing *I* with his own proper name, *John Perry*, can uniquely identify the right person, but only under the further crucial assumption that he would accept the statement ‘...and I am John Perry’— which smuggles the indexical back in. Even the deictic pronoun in 9d, uttered while pointing to the reflection of a shopper in the mirror, is not an adequate substitute. This utterance fails to express the desired meaning, unless it contains an implicit declaration that ‘the reflection is of me’, once again introducing a covert first person indexical.

This special mode of presentation of an utterance containing a first person indexical, whether overt or covert, is called *self-ascription* or *reference de se* in the philosophical literature. First person pronouns appear to be unique among linguistic expressions, in that they carry the force of speaker self-ascription in their inherent conventional semantics. A non-first-person expression such as a definite description (as in 9b), proper name (as in 9c), or third person deictic pronoun (as in 9d) can be used for self-ascription only if the speaker knows that the expression refers to him. Obviously the addressee must know this as well, if he is to interpret the utterance as a self-ascription by the speaker. In sum, first person pronouns, by virtue of their inherent linguistic content, force the speaker to self-ascribe, while apparently no other linguistic expressions have this effect.^v

Most work on self-ascription has focused on the first person, but second person pronouns have exactly the same self-ascriptive force, only applied to the *addressee* instead of the speaker.^{vi} Suppose another shopper witnesses the sugar spilling and she

wants to let Perry know that he is the culprit. If she wants to be certain that Perry understands, she must use a second person pronoun, as in 10a. An utterance of 10b-d will not guarantee that interpretation, unless accompanied by certain crucial background assumptions.

- (10) a. You are making a mess.
 b. The shopper with the torn sack is making a mess.
 c. John Perry is making a mess.
 d. He [*pointing to a reflection of John Perry in the mirror*] is making a mess.

A second person utterance like 10a, if taken to heart by the addressee, has the effect of inducing a self-ascription in the addressee's mind. For reasons exactly paralleling the ones given above for the first person utterance, none of the other locutions in 9b-d, repeated here as 10b-d, will necessarily convey the crucial information that will lead him to repair the torn sack in his shopping cart. For example, 10c works only if the addressee, upon hearing the statement '...and you are John Perry', would accept it as true. Again, this smuggles in a covert indexical. The second person pronoun, like the first person, is an essential indexical.

3.2 A solution to the problem

The sugar-spilling story illustrates a more general semantic problem: an agent can use and comprehend two different descriptions, or even different directly referential expressions such as pointing— without realizing that they refer to the same individual.

The fact that an agent can competently use referring expressions without being aware that they corefer leads to a family of semantic puzzles including, most famously, problems stemming from the failure of substitution of coreferential expressions within belief reports (see below). In contrast to this sea of referential uncertainty, indexical first and second person pronouns stand as islands of referential uniformity. For a given psychologically normal person, every use of the first person indexical pronoun is consistently understood to be referring to the same individual, regardless of the background context or assumptions. Similarly, every use of the second person indexical pronoun addressed to such a person is understood by him/her to be referring to the same individual.^{vii} Let us consider how to capture that difference between indexicals and other referring expressions.

Asher (1986, 128) observed that a framework for solving this class of problems was already ‘implicit in Discourse Representation Theory’ (DRT) as developed by Hans Kamp (1981) and Irene Heim (1982). A Discourse Representation Structure consists of a set of *reference markers* and a set of *conditions* on the anchoring of those reference markers. Asher (1986, 129) noted that reference markers can be used to model elements ‘in the mind of the recipient’ (i.e. the interpreter) of a speech act:

From the perspective of the theory of communication, reference markers are “conceptual individuals” that stand for real objects (with their properties represented by conditions) in the mind of the recipient. Semantically, reference markers are “pegs” on which the recipient can hang property ascriptions; conditions are those property ascriptions.

Similar to reference markers are the *file cards* posited by Heim (1982, 274ff), and the *notions* in the theory of Crimmins and Perry (1989) and Crimmins (1992).

A framework where reference markers (or file cards, notions, etc.) model conceptual individuals in the minds of speakers permits the following account of the crucial difference between indexicals and other referential expressions. In general two different reference markers in the mind of a single agent can, unbeknownst to the agent, map onto the same external object in the model. But a person indexical fixes a single reference marker: within the mind of a given agent, all indexical uses of the first person pronoun (when the agent is the speaker) and the second person pronoun (when the agent is the addressee) refer *via the same reference marker*. This is explained in more detail below.

Instead of *reference marker*, here I will adopt the term *notion* from the later work by Crimmins and Perry (1989) and Crimmins (1992), because the latter terminology evokes the private, mentalistic nature of what we are modeling, which will be important for what follows. In Crimmins and Perry's terminology, an agent's *notions* are constituents of his or her *beliefs*, which they describe as 'things in the head', 'concrete particulars that belong to agents just like arms, headaches, and bouts of the flu' (Crimmins and Perry 1989:688). The contents of beliefs, however, are *propositions*, which are public, abstract classifications of circumstances in the world. (Asher (1986) uses capitalized *Belief* for the 'object of belief' and lower-case *belief* for the mental state.) When Ann and Bob 'believe the same thing' it means Ann and Bob have (necessarily distinct) beliefs with the same propositional content. A belief's content

refers to the state of affairs in the public model of the world which holds iff the belief is correct.

Although I adopt the terminology from Crimmins and Perry (1989) and Crimmins (1992), any of the three traditions mentioned above is suited for our purposes. All of them explicitly assume that this *notion* (*reference marker*, *file card*, etc.) is part of an agent's mental representation of the world. This is shown for example by Asher's comment, quoted above, that 'reference markers are "conceptual individuals" that stand for real objects... in the mind of the recipient.' Hence nothing here should be construed as an endorsement of any one of these theories over the others, or indeed over any framework for modeling mental representations and their relations to the world that they represent (other examples include Fauconnier 1985; Jackendoff 1990; and van Ditmarsch, van der Hoek, and Kooi 2006; Cooper and Ginzburg 1996). The specific proposal presented in the following section is tailored to the needs of the topic addressed here, but the requisite ontological and theoretical assumptions are very modest and various frameworks are probably equally suitable for our purposes.

In DRT, reference markers are mapped onto objects in the domain of the model by a function called an *external anchor*. The corresponding function in Crimmins (1992), which maps *notions* to the public objects they are notions of, is called *ContentOf*; that term will be used here.^{viii} Notions are notated with the italic letter *n* subscripted with a mnemonic to distinguish them from each other. Since they are private, concrete, cognitive particulars, I will sometimes indicate the identity of the agent to whom the notion belongs with a prefixed superscript. For example, Barack Obama's notion of the city of Chicago might be notated ^{b.o.}*n*_{chicago}; his notion of Hillary Clinton might be ^{b.o.}*n*_{h.c.}.

Then $ContentOf^{(b.o.)}n_{chicago} = Chicago$, and $ContentOf^{(b.o.)}n_{h.c.} = Hillary\ Clinton$, where ‘Chicago’ and ‘Hillary Clinton’ are individuals in the public model of the world. Although the prefixed superscript is sometimes omitted for perspicuity, every notion is necessarily a private, concrete, cognitive particular in the mind of a single agent.

As noted already, a single agent can sometimes have two distinct notions of the same individual. Consider Kripke’s (1979) famous story of Pierre, which was analyzed by both Asher (1986) and Crimmins (1992) in their respective frameworks. In Kripke’s (1979) story, Pierre has two different notions of the city of London, one acquired by reading French tourist brochures, which he associates with the name ‘Londres’ and believes to be pretty (call it $^{Pierre}n_{Londres}$), and one acquired by visiting London, which he associates with the name ‘London’ and does not believe to be pretty (call it $^{Pierre}n_{London}$). Pierre is unaware that his ‘Londres’ and his ‘London’ are actually the same city. To use the present terminology, he is unaware that his two *notions* have the same content: $ContentOf^{(Pierre)}n_{Londres} = London$, and $ContentOf^{(Pierre)}n_{London} = London$. Pierre ascribes prettiness to London via one of those notions but not the other.

The semantic puzzle is how the following two belief reports could both be true, despite the fact that ‘Londres is pretty’ and ‘London is pretty’ express the same proposition, and Pierre does not hold irrational or contradictory beliefs:

- (11) a. Pierre believes that Londres is pretty.
 b. Pierre does not believe that London is pretty.

The answer offered by Asher, Crimmins and others, is that beliefs are distinguished by more than just their propositional content. The two sentences report two *different* beliefs with the *same* propositional content. Pierre has one of those beliefs but lacks the other.

Much as the mental representation of an *individual* is called a *notion* (abbreviated *n*), the mental representation of a *relation* is called an *idea* (abbreviated *i*). A particular *belief structure* in the mind of an agent is made up of certain *ideas* and *notions*. For example, the structure of the belief reported in 11a contains Pierre’s *idea* of the ‘pretty’ property ($^{Pierre}i_{pretty}$) and his *notion* of London ($^{Pierre}n_{Londres}$)— specifically, the notion that was acquired by reading French tourist brochures. For Crimmins and Perry (1989) and Crimmins (1992) a *belief structure* contains a *k*-ary *idea* and a sequence of *k* *notions* (in this example, $k = 1$), which we will notate with angle brackets, as in 12a, following Crimmins (1992). The propositional content of that belief contains a *k*-place *relation* and *k* *individuals* filling the argument roles of that relation, as in 12b, which is represented in a more standard logical notation. The latter proposition is the content of the former belief structure, as indicated in 12c.

(12) a. $\langle i_{pretty}; n_{Londres} \rangle$

b. **pretty**’(London)

c. $ContentOf(\langle i_{pretty}; n_{Londres} \rangle) = \mathbf{pretty}'(\text{London})$

A basic relation *Believes* holds between an agent *A*, a belief structure β , and a proposition *p* (the propositional content of β):^{ix}

(13) $\text{Believes}(A, \beta, p)$

The belief reports in 11a, b are represented in 14a,b respectively. Note that they have the same propositional content:

(14) a. $\text{Believes}(\textit{Pierre}, \langle i_{\textit{pretty}}, n_{\textit{Londres}} \rangle, \textit{pretty}'(\textit{London}))$

b. $\neg \text{Believes}(\textit{Pierre}, \langle i_{\textit{pretty}}, n_{\textit{London}} \rangle, \textit{pretty}'(\textit{London}))$

Unless Pierre knows that his two notions are of the same city, his beliefs and other mental attitudes involving one of them will remain disconnected from those involving the other, within his mind.

Turning now to the problem of the essential indexical, let us posit that for a given speaker, all uses of the first person pronoun refer via the same notion. Crimmins (1992:163-5) calls this special notion a *self-notion*, and uses it to address the problem of the essential indexical. An agent who self-ascribes a property has ‘a belief about herself *via her self-notion*. ... There is no special indexical *object of belief*, but there is a special kind of notion—the self-notion—that each of us has.’ (Crimmins 1992:164-5) Similarly, in discussing an elaboration of Kripke’s story, Asher (1986, 146) posits ‘a special, “irreducibly indexical” reference marker ‘*i*’ to denote Pierre to himself.’

We axiomatize the ‘irreducibly indexical’ character of the self-notion by positing that an agent x ’s self-notion ${}^x n_{\textit{self}}$ is necessarily a notion of x :

(15) *The Self-Notion Axiom*. Necessarily, $\forall x[\textit{ContentOf}({}^x n_{\textit{self}}) = x]$.

Self-ascription is simply ascription via the self-notion. For any agent A , A 's normal notion of A is the self-notion. It is the notion involved in the vast majority of beliefs about A that A holds. In addition to beliefs, the self-notion is also common in other representational mental states such as intentions, desires, and perceptions. If the content of such a mental state includes the agent A herself, then the usual notion of A employed by that agent within the structure of her mental state is the self-notion (Searle 1983). However, the self-notion is not necessarily involved in every belief about A that A holds: in the early part of Perry's sugar-spilling story, before his epiphany, A (=Perry) believes that A is spilling sugar via *other* notions of A : the notion of 'the shopper with a torn sack', the notion of a person named John Perry, or the notion of the person reflected in the mirror. Each of these notions that he has is a notion of himself, but he is unaware of that fact until his epiphany. The new belief he acquires, that explains why he fixes the sugar sack in his cart, is crucially a belief via his 'irreducibly indexical' self-notion.

The pronoun I is grammatically specified for a *speaker* to refer via her self-notion; the pronoun you is grammatically specified for an *addressee* to interpret via her self-notion. In Perry's story, his pronoun I in 9a is grammatically specified for referring via his self-notion $^{perry}n_{self}$. (In contrast, 9c, for example, says only that someone named 'John Perry' is making a mess.) Perry's self-notion $^{perry}n_{self}$ is the same notion involved in countless other self-ascriptions by him that make up normal conscious awareness: that he is pushing his shopping cart, that the hand at the end of his arm will move when he wills it to do so, and so on. The Self-Notion Axiom allows him to connect all those beliefs with the same individual. This is why the first person utterance explains his fixing

the sugar sack in his cart. But other locutions such as those in 9b-d do not necessarily involve the self-notion. Utterance 9c, for example, involves the notion of someone named John Perry; call it ${}^{perry}n_{named-JohnPerry}$. The Self-Notion Axiom does not apply here, so he must rely on background knowledge instead. Only if he knows that his notion ${}^{perry}n_{named-JohnPerry}$ and his self-notion ${}^{perry}n_{self}$ have the same content, will his utterance of 9c express a belief that explains why he fixes his sugar sack.

In short, the indexical first and second person pronouns *require* reference via the self-notion. Other referring expressions merely *allow* reference via the self-notion, and only if the appropriate background assumption is made. Because first person pronouns require reference via the self-notion, such reference is effectively paraphrased using a first person pronoun. Thus the necessary background assumption for the use of a non-indexical is the one described in Section 3.1 above as ‘smuggling the indexical back in’: regarding 9c, for example, it is the assumption that he would accept the statement ‘...and I am John Perry’. This analysis solves the problem of the essential indexical.

4. Self-ascription in the grammar

4.1 Indexical pronouns

Now let us sketch a grammatical framework for natural language utterances, including indexical pronouns. In this theory the compositional semantics of a natural language grammar builds the *conceptual structures* of agents rather than just building the propositional contents (Jackendoff 1990, inter alia). Conceptual structures are just a generalization of the belief structures discussed above. The *Believes* relation that we borrowed from Crimmins and Perry’s account of belief reports is too specific for our

purposes: it presumes speaker sincerity and addressee credulousness. Communication can take place between insincere speakers and skeptical addressees, and there are other illocutionary acts besides assertion (such as questions and imperatives). So we abstract away from the particular intentional modes such as belief, desire, doubt, and so on, replacing the *Believes* relation with the more general *Conceives* relation, and replacing the belief structure with a conceptual structure. Like the belief structure, the conceptual structure is understood here as a private, concrete, cognitive particular.

We posit a translation function V from expressions of a natural language to conceptual structure constituents such as notions, ideas, and complex entities composed from them. Since those structures are concrete cognitive particulars in the minds of particular agents, V must also be relativized to agents. So the function V (for English) takes an ordered pair as its argument, the first element ranging over (English-competent) agents and the second over expressions of the language (English). V returns constituents of the agent's conceptual structure. For example, suppose Mary and Paula, both speakers of English, know Bill by the name [*Bill*], and know the city of Austin, Texas by the name [*Austin*]. Square brackets indicate linguistic expressions ([*Bill*], [*Austin*], etc.); mental notions are represented as above (${}^m n_{Bill}$ for Mary's notion of Bill, etc.); normal font indicates elements of the public model of the world (Bill, Austin, etc.):

(16) Example of the English translation function V .

$$V(\langle \text{Mary}, [\text{Bill}]^{S,A} \rangle) = {}^m n_{\text{Bill}}$$

$$V(\langle \text{Mary}, [\text{Austin}]^{S,A} \rangle) = {}^m n_{\text{Austin}}$$

$$V(\langle \text{Paula}, [\text{Bill}]^{S,A} \rangle) = {}^p n_{\text{Bill}}$$

$$V(\langle \text{Paula}, [\text{Austin}]^{S,A} \rangle) = {}^p n_{\text{Austin}}$$

(17) The function *ContentOf*

$$\text{ContentOf}({}^m n_{\text{Bill}}) = \text{Bill}$$

$$\text{ContentOf}({}^m n_{\text{Austin}}) = \text{Austin}$$

$$\text{ContentOf}({}^p n_{\text{Bill}}) = \text{Bill}$$

$$\text{ContentOf}({}^p n_{\text{Austin}}) = \text{Austin}$$

Contextual coordinates A (the set of addressees) and S (the set of speakers, a singleton set unless multiple speakers are speaking in unison) are shown for completeness, but do not affect the output of the function V for these particular words: Mary and Paula understand these words in the same way regardless of who the interlocutors are. The transitive verb [*likes*] translates as a function from pairs of notions to conceptual structures:

$$(18) \text{ a. } V(\langle \text{Mary}, [\text{likes}] \rangle) = \lambda y \lambda x \langle {}^m i_{\text{likes}}; x, y \rangle$$

$$\text{ b. } V(\langle \text{Paula}, [\text{likes}] \rangle) = \lambda y \lambda x \langle {}^p i_{\text{likes}}; x, y \rangle$$

Although the details are unimportant, assume for concreteness that the conceptual structure induced by a transitive sentence is built up in the syntax by function application,

where the object NP is fed to the function first, and then the subject NP, in the familiar manner (except that it derives conceptual structures instead of semantic contents).

For example, suppose Mary says to Paula, ‘Bill likes Austin.’ Then the speaker and addressee conceptual structures are calculated to be the following:

- (19) [Bill likes Austin.]^{{Mary},{Paula}} (Mary speaking to Paula)
- a. speaker: Conceives(*Mary*, $\langle {}^m i_{likes}; {}^m n_{Bill}, {}^m n_{Austin} \rangle, \dots$)
- b. addressee: Conceives(*Paula*, $\langle {}^p i_{likes}; {}^p n_{Bill}, {}^p n_{Austin} \rangle, \dots$)

We calculate the propositional contents of these beliefs by applying the function *ContentOf* (see 17) to the respective ideas and notions:

- (20) [Bill likes Austin.]^{{Mary},{Paula}} (Mary speaking to Paula)
- a. speaker: Conceives(*Mary*, $\langle {}^m i_{likes}; {}^m n_{Bill}, {}^m n_{Austin} \rangle, \mathbf{likes}'(\text{Bill}, \text{Austin})$)
- b. addressee: Conceives(*Paula*, $\langle {}^p i_{likes}; {}^p n_{Bill}, {}^p n_{Austin} \rangle, \mathbf{likes}'(\text{Bill}, \text{Austin})$)

Summarizing so far, the representations above indicate certain information about the utterance of *Bill likes Austin* by Paula to Mary. They show the conceptual structures associated with that sentence, within Paula’s and Mary’s respective minds; and they show the propositional contents of those conceptual structures.

Turning to the indexical pronouns *you* and *I*, we argued above that these pronouns are self-ascriptive. Hence they translate as self-notions. I furthermore propose that *self-ascription exhausts the person semantics of these forms* (the *Self-Ascription Monopoly*;

see §4.2 below). These pronouns indicate self-ascription but there is no additional specification that they must ‘refer to’ or ‘be anchored to’ the addressee and speaker. Let us see how this works.

For every speaker, *I* translates as a self-notion, and for every addressee, *you* translates as a self-notion:

(21) Translations of *I* and singular *you* (preliminary).

a. $\forall s \in S [V(\langle s, [I]^{S,A} \rangle) = {}^s n_{self}]; \forall x \notin S [V(x, [I]^{S,A}) \text{ is undefined}]$

b. $\forall a \in A [V(\langle a, [you]^{S,A} \rangle) = {}^a n_{self}]; \forall x \notin A [V(x, [you]^{S,A}) \text{ is undefined}]$

Crucially, V is a partial function. When it applies to the first person pronoun [*I*], then V is defined only if the first argument is a speaker of the utterance. For the pronoun [*you*], the first argument must be an addressee. For any other agents, V is undefined for these words.

For example, consider the sentence *I like Austin*, uttered by Mary to Paula. Hence the set of speakers $S = \{Mary\}$ and the set of addressees $A = \{Paula\}$, as shown.

(22) [*I like Austin*]^{{Mary},{Paula}} (Mary speaking to Paula)

a. speaker: Conceives(*Mary*, $\langle {}^m i_{likes}; {}^m n_{self}, {}^m n_{Austin} \rangle$, **likes'**(*Mary*, *Austin*))

b. addressee: Conceives(*Paula*, $\langle {}^p i_{likes}; \boxed{\eta}, {}^p n_{Austin} \rangle$, **likes'**($\boxed{\chi}$, *Austin*))

V returns Mary’s self-notion (${}^m n_{self}$) in the first argument place of her ${}^m i_{likes}$ idea. But V does not provide a notion in the first place of Paula’s ${}^p i_{likes}$ idea, so the place-holder η

appears. Since no notion is provided, no content can be calculated for the position indicated by the place-holder, χ .

As addressee, Paula (and anyone else who is interested, such as an eavesdropper) solves for χ and η in the following way. She builds a mental model of Mary's belief state and derives the content of that belief, much as we, as linguists, have done in our analysis: by translating the pronoun *I* (using rule 21a). Mary is the speaker so it translates as Mary's self-notion within her conceptual structure. By the *Self-Notion Axiom* (15), the content of Mary's self-notion is just Mary. Paula seeks to understand Mary's intended message by making her content the same as Mary's content. The first role of the **likes'** relation in Mary's content is filled by 'Mary', so 'Mary' fills that role in Paula's content as well. Hence $\chi = \text{Mary}$. Paula solves for η by finding a notion of her own with Mary as its content, hence ${}^P n_{\text{Mary}}$. This completes the derivation:

- (23) [I like Austin]^{{Mary}, {Paula}} (Mary speaking to Paula)
- a. speaker: Conceives(Mary, $\langle {}^m i_{\text{likes}}, {}^m n_{\text{self}}, {}^m n_{\text{Austin}} \rangle$, **likes'**(Mary, Austin))
- b. addressee: Conceives(Paula, $\langle {}^m i_{\text{likes}}, \boxed{{}^P n_{\text{Mary}}}, {}^P n_{\text{Austin}} \rangle$, **likes'**($\boxed{\text{Mary}}$, Austin))

Second person pronouns are analyzed similarly, but for self-ascription by the addressee. As in the previous example, the first step of the following derivation gives the grammatically specified self-ascriptions (24), and in the second step the other interlocutor fills in remaining values (25), as explained above:

- (24) [I like you]^{{Mary},{Paula}} (Mary speaking to Paula)
- a. speaker: Conceives(*Mary*, $\langle {}^m i_{likes}; {}^m n_{self}; \boxed{\eta_1} \rangle$, likes'(*Mary*, $\boxed{\chi_1}$))
- b. addressee: Conceives(*Paula*, $\langle {}^p i_{likes}; \boxed{\eta_2}; {}^p n_{self} \rangle$, likes'($\boxed{\chi_2}$, *Paula*))

- (25) [I like you]^{{Mary},{Paula}} (Mary speaking to Paula)
- a. speaker: Conceives(*Mary*, $\langle {}^m i_{likes}; {}^m n_{self}; \boxed{{}^m n_{Paula}} \rangle$, likes'(*Mary*, \boxed{Paula}))
- b. addressee: Conceives(*Paula*, $\langle {}^p i_{likes}; \boxed{{}^p n_{Mary}}; {}^p n_{self} \rangle$, likes'(\boxed{Mary} , *Paula*))

Summarizing, the translation function for a natural language builds conceptual structures in the minds of linguistically competent agents. Since those beliefs are agent-specific, the function is relativized to the agent. For most English expressions it provides values (to those agents familiar with the expression). But for indexical personal pronouns it provides values only to the self-ascriber: to the speakers of a first person pronoun and the addressees of a second person pronoun. That value is a self-notion, which captures the phenomenon of self-ascription and solves the problem of the essential indexical. The translation function does not provide a value for first and second person pronouns to agents other than the designated self-ascriber, so those others must solve for the relevant values by building a model of the self-ascriber's mental state.

Note that this latter process requires the human cognitive ability known as *Theory of Mind*, the ability to impute mental states to others and draw inferences from those mental states (Premack and Woodruff 1978). Important convergent evidence for this theory is thus provided by the fact that children at an age before theory of mind is fully mastered have special difficulty precisely with non-self-ascribed pronouns, that is, with

the speaker-production of second person pronouns and addressee-comprehension of first person pronouns. See Section 6 below.

4.2 The Self-Ascription Monopoly

While it has long been observed that the interpretation of indexical pronouns involves self-ascription (Hector-Neri Castañeda 1977; Kaplan 1977; Perry 1979; Kripke 1979; Lewis 1979; Richard 1983), the present proposal goes a step further by positing that *all pronominal reference to speech act participants takes place via self-ascription*. The phrase *via self-ascription* includes *direct* pronoun interpretation by its self-ascriber (speaker for first person, addressee for second person), as provided by the translation function; as well as *indirect* pronoun interpretation by a non-self-ascriber, who makes an inference from another interlocutor's self-ascription (addressee for first person, speaker for second person). The *Self-Ascription Monopoly* is our name for that foundational assumption that self-ascription holds a monopoly on pronominal reference to speech act participants:

- (26) *The Self-Ascription Monopoly*: Only as a consequence of grammatically specified self-ascription can a pronoun be knowingly used to refer to a speaker or addressee.

So far our only example of 'grammatically specified self-ascription' is the self-ascription specified by the provisional translation rules (21) for first and second person singular

English pronouns. The next section generalizes the account to plural pronouns, where the *Self-Ascription Monopoly* will do more work for us.

4.3 Plural pronouns

The goal of this section is to present a *de se* semantics of first and second person plural pronouns, thus extending the theory sketched in the previous section to plurals. The semantics to be proposed, basically a self-ascriptive variant of an associative plural, is a stipulation at this stage in our presentation. Some linguistic evidence for the present account is presented in Section 4.4. In Section 5 below I address the larger goal of showing that the *de se* theory makes regular plural semantics of plural pronouns impossible in principle, which thus predicts the Associative Plural Generalization.

Recall from Section 2 above that first person plural pronouns (*we, us, our, ourselves*) can be used to refer to any group of individuals that *includes* the speaker or speakers, while a second person plural pronoun can be used for reference to any group that *includes* all the addressees.

(27) Possible interpretations of English first person singular and plural pronouns

<u>interpretation</u>	<u>example</u>
a. 'speakers'	sg: <i>I need a drink.</i> pl: <i>We are the champions!</i> (unison)
b. 'speakers + others'	<i>We want you to come to dinner.</i>
c. 'speakers + addressees'	<i>Shall we go?</i>
d. 'speakers + addressees + others'	<i>Can't we all get along?</i>

(28) Possible interpretations of English second person forms.

- | | | |
|-------------------------|-----|--|
| a. ‘addressees’ | sg: | <i>You should behave yourself.</i> |
| | pl: | <i>You should behave yourselves.</i> |
| b. ‘addressees + other’ | | <i>How do you guys handle promotions over in Philosophy?</i> |

Recall further that many languages distinguish first person plural inclusive pronouns, which include the addressees, from exclusive pronouns, which exclude them:

(29) Possible interpretations of first person plural.

<u>interpretation</u>	<u>example (Indonesian)</u>
a. ‘multiple speakers’	<i>kami</i>
b. ‘speaker + other’	<i>kami</i>
c. ‘speaker + addressee’	<i>kita</i>
d. ‘speaker + addressee + other’	<i>kita</i>

Now let us spell out the semantics of first and second person plural pronouns.

To a first approximation, a first person pronoun is interpreted as a set X of individuals such that every speaker self-ascribes membership in X , while a second person pronoun is interpreted as a set X such that every addressee self-ascribes membership in X . However, this characterization in terms of sets is a simplification, given the semantics of plural nominals more generally. When uttering the plural sentence in 27a (*We are the*

champions!), for example, the speaker does not predicate championhood of himself but rather predicates it collectively of his team as a whole. (Thus the speaker need not agree to the assertion *I am a champion!*) The pronoun *we* refers to the team, and the speaker self-ascribes membership in the team and asserts that the team is a champion team.

Collective predication has been analyzed in various ways involving slightly different ontological commitments, but the analysis of the semantics of plurality is essentially independent of the issues addressed here, as we will show. The semantics of plurals is concerned with interpreting a plural nominal, here a plural pronoun, and semantically composing it with the rest of the sentence. As in the example in the previous paragraph, all that we need is a way to relate the collective entity denoted by a plural pronoun to the set of individuals (actually individual *atoms*; see below) that constitute that collection. The present claim is that the person feature is interpreted as follows: the speaker (of a first person pronoun) or addressee (of a second person pronoun) self-ascribes membership in that set of individuals.

For example, Link (1998) defined *sums* (or *plural objects*) as individuals which are aggregates of other individuals. Hence if a and b denote two atoms, then $a \oplus b$, the *individual sum* of a and b , is a third individual. Sums are partially ordered by a 2-place part relation Π satisfying the biconditional

$$(30) \quad a \Pi b \Leftrightarrow a \oplus b = b.$$

The part relation is transitive and forms a semi-lattice. So if a , b , and c denote three atoms, then a is part of $a \oplus b$; $a \oplus b$ is part of $a \oplus b \oplus c$; a is part of $a \oplus c$; $a \oplus c$ is part

of $a \oplus b \oplus c$; and a , b , and c are each parts of $a \oplus b \oplus c$. Minimal elements, those which have only themselves as parts, are called *atoms*; in this example the atoms are a , b , and c . For some entity d , $\mathbf{AT}(d)$ designates the set of atoms within d . Hence $\mathbf{AT}(a \oplus b \oplus c) = \{a, b, c\}$. For convenience, below we will call this set of atoms within the referent of some expression the *reference set* of that expression.

Under the view sketched above, a collective predicate such as *are the champions* could be predicated of $a \oplus b$ without being predicated of a or b . Now, returning to indexical pronouns, we will say that a first person pronoun denotes an entity x such that every speaker self-ascribes membership in the set $\mathbf{AT}(x)$. A second person pronoun denotes an entity x such that every addressee self-ascribes membership in the set $\mathbf{AT}(x)$. The analysis sketched above applies equally to the plural pronoun *we* in the collective predication in 27a (*We are the champions!*). The predicate *are the champions* applies to the group but not to the individuals making up that group.

In contrast to collective predicates, distributive predicates such as *have blue eyes* must distribute over the members, so that 31a entails 31b (Landman 2000, 118-9, citing Scha 1981) Scha (1981) proposed the meaning postulate in 31c to get the correct interpretation.

- (31) a. The boys have blue eyes.
 b. Every boy has blue eyes.
 c. $\text{have.blue.eyes}(x)$ iff $\forall a \in \mathbf{AT}(x): \text{have.blue.eyes}(a)$

Applying Scha's meaning postulate, sentence 32 is also interpreted distributively, just like 31a.

(32) We have blue eyes.

The speaker self-ascribes membership in a collection x and predicates *have blue eyes* of x . By the meaning postulate 31c, the sentence entails that each member of the group, including the speaker of course, has blue eyes.

It is the self-ascription connecting the speaker or addressee to the denotation of the pronoun that is the topic of the present paper. The interpretation of that pronoun within the sentence is a topic for the semantics of plurals, if the pronoun is plural.

To capture personal pronoun *systems* we posit two privative person features, [*spk*] and [*addr*] ('speaker' and 'addressee'), and a number (NUM) feature with values ranging over singular, plural, dual, and so on, depending on the language.

(33) Semantic interpretation. A personal pronoun bearing the multivalued feature [NUM] and one or more of the privative features [*spk*] and [*addr*] denotes an entity G . Let $\mathbf{AT}(G) = A$. Then A is a set of individuals constrained by the features [NUM], [*spk*] and [*addr*] as follows:

a. [NUM *sg*]: $|A| = 1$

[NUM *pl*]: $|A| > 1$

[NUM *dual*]: $|A| = 2$

...etc.

b. [*spk*]: Every speaker self-ascribes membership in A .

c. [*addr*]: Every addressee self-ascribes membership in A .

Following are translations for pronouns bearing the person features [*spk*] and [*addr*]. Note that the speaker and addressee indices S and A range over *sets* of individuals. A pronoun marked with both features [*spk*, *addr*] is subject to both conditions, hence an inclusive pronoun.

(34) Constraints imposed by the pronoun features [*spk*] and [*addr*] on the assignment function V as applied to a pronoun Pron:

a. $\forall s \in S [V(s, \text{Pron}[\textit{spk}]^{S,A}) = x \wedge {}^s n_{self} \in \mathbf{AT}(x)];$

$\forall x \notin S [V(x, \text{Pron}[\textit{spk}]^{S,A}) \text{ is undefined}]$

b. $\forall a \in A [V(a, \text{Pron}[\textit{addr}]^{S,A}) = x \wedge {}^s n_{self} \in \mathbf{AT}(x)];$

$\forall x \notin A [V(x, \text{Pron}[\textit{addr}]^{S,A}) \text{ is undefined}]$

In prose, the assignment function for a speaker and a pronoun marked [*spk*] assigns it to some entity whose constituent atoms include the speaker's self-notion (${}^s n_{self}$). The assignment function for an addressee and a pronoun marked [*addr*] assigns it to some entity whose constituent atoms include the addressee's self-notion (${}^a n_{self}$). The universal quantifier indicates that if there are multiple speakers or hearers, then that translation applies to each of them. If Mary and Susan say *We are the champions!* in unison, then Mary's reference set for *we* includes Mary's self-notion, and Susan's reference set for *we* includes Susan's self-notion. The rules in 34 apply to singular and plural pronouns across languages, replacing the provisional ones in 21 for English singular pronouns.

Summarizing, speakers (/addressees) self-ascribe membership in the collection denoted by a [*spk*] (/ [*addr*]) pronoun. Inclusive pronouns such as Indonesian *kita* (recall 29c,d) are marked with both features, [*spk, addr*], so both conditions apply: every speaker and every addressee self-ascribes membership in the reference set (35d). English *we* is 'general', subsuming both inclusive and exclusive, so the [*addr*] feature is optional (see 35e).

(35) Pronoun features.	<u>NUM</u>	<u>person</u>
a. <i>1st singular</i> (e.g. English <i>I</i>):	[<i>sg</i>]	[<i>spk</i>]
b. <i>2nd singular</i> (e.g. English <i>yourself</i>):	[<i>sg</i>]	[<i>addr</i>]
c. <i>1st plural exclusive</i> (e.g. Indonesian <i>kami</i>):	[<i>pl</i>]	[<i>spk</i>]
d. <i>1st plural inclusive</i> (e.g. Indonesian <i>kita</i>):	[<i>pl</i>]	[<i>spk, addr</i>]
e. <i>1st plural general</i> (e.g. English <i>we</i>):	[<i>pl</i>]	[<i>spk, (addr)</i>]
f. <i>2nd plural</i> (e.g. English <i>y'all</i>):	[<i>pl</i>]	[<i>addr</i>]
g. <i>3rd singular</i> (e.g. English <i>he</i>):	[<i>sg</i>]	
h. <i>3rd plural</i> (e.g. English <i>they</i>):	[<i>pl</i>]	

We can now define ‘grammatically specified’ as ‘grammatically encoded by the privative features [*spk*] and [*addr*]’, for the purpose of the *Self-Ascription Monopoly* in 26 above (‘Only as a consequence of *grammatically specified* self-ascription can a pronoun be knowingly used to refer to a speaker or addressee.’) The *Self-Ascription Monopoly* was introduced in order to guarantee self-ascription for indexical pronouns: a speaker uttering a first person pronoun *must* self-ascribe; and even an addressee hearing a first person pronoun must infer that the speaker has self-ascribed. That helped us to solve the problem of the essential indexical.

The *Self-Ascription Monopoly* does quite a bit more work for us now. It immediately captures the following *exclusions* from membership in the reference sets of pronouns:

(36) Further consequences of the *Self-Ascription Monopoly*.

- a. Third person pronouns normally exclude speakers and addressees. For example, the reference set of *they* in *They are ready* normally excludes any speaker or addressee. This follows since *they* lacks either of the features [*spk*] and [*addr*].
- b. Second person pronouns exclude speakers. For example, the reference set of *you* in *You behaved yourselves* cannot include any speaker. This follows since second person pronouns lack the [*spk*] feature.
- c. First person exclusive pronouns exclude addressees. This follows since exclusive pronouns lack the [*addr*] feature.

The proviso ‘knowingly’ in the statement of the Self-Ascription Monopoly allows for accidental reference to the speaker or addressee. Recall for example 9d above, where the speaker points to a reflection in the mirror and says *He is making a mess*. Since he does not know that the reflection is of himself, he uses a third person instead of a first person pronoun.

The Self-Ascription Monopoly places a condition on pronouns but not on other noun phrases. Sometimes a phrase like *this author* refers to the speaker (or writer), and a third person title like *your honor* can refer to the addressee (Siewierska 2004). So this principle may be partly explained as a markedness effect of the paradigmatic opposition between pronoun forms. The first person feature [*spk*] designates self-ascription by the speaker, so that form is selected over any pronoun forms lacking that feature, for reference to the speaker. The second person feature [*addr*] designates self-ascription by

the addressee, so that form is selected over any other pronoun form for reference to the addressee. For example, an Optimality Theoretic faithfulness constraint might have this effect.

But the *Self-Ascription Monopoly* could also have a deeper explanation, rooted not in the personal pronoun paradigm but rather in the paradigmatic opposition between reference *de se* and reference *de dicto*. As discussed above, reference *de se* expresses self-knowledge. A *de se* attribution expresses a ‘self-locating belief’, to use Perry’s (1979) term. A person who is unable to self-ascribe any properties is psychologically lost. Merely ascribing a property to oneself, without self-ascribing it, suggests confusion or ignorance (cp. Grice’s 1975 maxim of quantity).

4.4 Self-ascription *by* versus reference *to*

In the present *de se* analysis of indexicals, a first (/second) person pronoun indicates self-ascription *by* every speaker (/addressee). This differs from the standard view, according to which the first and second person are grammatically specified for reference *to* the speakers and addressees. These two hypotheses will be teased apart on the basis of utterances with multiple addressees. Imagine a teacher telling her class:

(37) Write your name at the top of the page.

Each addressee *x* interprets the second person pronouns *your* as referring to *x*. Even if Tommy and Mary are both addressees, Tommy understands the teacher as instructing him to write his own name, not Mary’s. This is exactly what the *de se* analysis predicts.

The second person pronoun is specified for self-ascription *by* each addressee. So each addressee x is being told write x 's name at the top of the page, not just to write the name of some addressee.

The same point can be made with plural pronouns. The leader of a workshop for married men (with no wives present) asks the men about their respective marriages:

(38) How often do you kiss each other?

Each addressee x interprets the second person pronoun as referring to a collection that includes x , in this case most plausibly the collection ' x and x 's wife.' As in the previous example, an addressee is not free to include just *any* addressee in the reference set of *you*; he must include himself. Again, this interpretation is correctly predicted by the de se theory presented here. The pronoun feature [*addr*] specifies that every addressee self-ascribes membership in the reference set G of the pronoun.

What is predicted by other theories of indexicals, such as double indexing approaches (Kamp 1971; David Kaplan 1977, *inter alia*)? According to a common analysis of indexical pronouns, the addressee in the utterance context gets assigned to a second person pronoun by the assignment function that assigns values to pronouns in a context. For example, Chierchia and McConnell-Ginet (2000, 342) posit the following character for *you*, where V is the assignment function, c is an utterance context, w is a world, i is an instant of time, and $\text{adr}(c)$ is the addressee in c : ^{x}

- (39) For any c , $V(\text{you})(c)$ is a function such that for any w and any i , $V(\text{you})(c)(\langle w, i \rangle) = \text{adr}(c)$.

That is, this assignment function assigns *you* to the addressee, in all worlds and times.

Chierchia and McConnell-Ginet do not analyze utterances with multiple addressees, but Schlenker (2003b; 2003a) does. Schlenker proposes that a second person pronoun carries the presupposition that its reference set ‘contains a hearer’ of the utterance (2003a, 5; 2003b, 110). A second person pronoun bears the feature +hearer*, which encodes the following presupposition, where s is his assignment function, c^* is an utterance context, and x_i is the variable associated with the pronoun (Schlenker 2003a, 5):

- (40) +hearer*(x_i, c^*) is true _{s} iff $s(x_i)$ contains a hearer of $s(c^*)$. Otherwise it is false _{s} .

An utterance with this feature is felicitous only if the presupposition in 40 is true, so the pronoun *you* is uttered felicitously only if the pronoun’s reference set includes a hearer.

Schlenker (2003b, 50) analyzes the following example with two addressees, in which ‘a demonstration typically completes the sentence’ (his examples 26 and 31a):

- (41) You₁ [pointing] are elected, but you₂ [pointing] are not.

Since *you* carries the presupposition that its reference set includes an addressee— and being singular, includes no one else— then as long as these two pronouns refer to distinct

individuals from the set of addressees, this sentence is not contradictory (Schlenker 2003a, 50-52; Schlenker 2003b, 4).

On the *de se* theory, all and only addressees self-ascribe when they hear a second person pronoun. In 41 the first clause is addressed to (the real world referent of) *you*₁ and the second to (the real world referent of) *you*₂, and each addressee self-ascribes his own clause. Evidence favoring the *de se* account comes from the fact that speaker eye gaze, which is known to signal addressee-hood, is a nearly mandatory accompaniment to the pointing for the second person cases like 41, but not so for third person (e.g. when giving directions). The reader can verify this by playing the role of speaker in 41. The shifting eye gaze and demonstration (pointing) distinguishes two different addressees, not just two different pronoun referents. When the addressee changes, the shifting reference follows automatically from the *de se* theory.

In any case, 37 and 38 above pose a different problem. Everyone hears the same pronoun and there is no ‘demonstration’ (pointing) to restrict its reference. Nothing in Schlenker’s analysis ensures that each addressee should self-ascribe the property in question, rather than ascribing it to some other addressee. If Tommy and Mary are both addressees of 37, then Tommy could interpret *you* as referring to Mary, and therefore conclude that his writing Mary’s name on the page would satisfy the teacher’s demand. Similarly, nothing would prevent one of the men hearing 38 from interpreting the plural *you* as referring to some set that excludes him, but includes a classmate. But those interpretations are unavailable.

One way to fix the standard theory would be to assume that the contextual addressee variable (also called an *index* or *coordinate*) is *identified* with the second

person pronoun's variable. More accurately, to allow for plural pronouns, the addressee index would be identified with a variable over elements of the pronoun's reference set. For example, assuming quantification into speech acts (on which, see Krifka (2001)), any quantifier binding the contextual addressee variable from outside the illocutionary operator will bind the pronoun variable as well. In 37 for example, the imperative illocutionary force distributes over the children: 'for each addressee x , the speaker commands x to write x 's name at the top of the page.' Similarly, in 38 the interrogative distributes over the addressees: 'for each addressee x , the speaker asks x how often [the people in some set that includes x] kiss each other.' When the assignment function assigns an extension to that variable, the addressee and the pronoun referent are inextricably linked, as desired.

Let us suppose that such a move, of identifying the addressee index with the pronoun variable, is successfully incorporated into a double indexing model. It would partially mimic the effects of assuming self-ascription by the addressees. But clearly this move is not a necessary consequence of a double indexing model. In contrast, the *de se* theory correctly captures these facts without any stipulation. The link between each addressee and the pronoun reference associated with her, as illustrated by the scenarios above, is a necessary consequence of the theory.

5. Explaining the Associative Plural Generalization

As we saw in Section 2, the Associative Plural Generalization raises the puzzling question of why first and second person pronouns have *associative* semantics rather than the *regular plural* semantics found with most plural nominals. Regular plural semantics

involves universal quantification over the members of the reference set, while associatives involve existential quantification over the members of the reference set:

(42) a. regular plural NPs: ‘Every member of the reference set is...’

(e.g. *the dogs*: ‘Every member of the reference set is a dog.’)

b. associative NPs: ‘Some member of the reference set is...’

(e.g. Hungarian *Péter-ék*: ‘Some member of the reference set is named *Péter*.’)

On the standard theory of indexicals, first and second person plurals are always specified for existential rather than universal quantification:

(43) First and second person plurals on the standard theory.

a. 1st person exclusive: ‘Some member of the reference set is the speaker.’

b. 1st person inclusive: ‘Some member of the reference set is the speaker; and
some member of the reference set is the addressee.’

c. 2nd person: ‘Some member of the reference set is the addressee.’

The question for the standard theory is why such pronouns always correspond to the relatively rare associative NPs 42b, and never to the very common regular plurals 42a. As discussed above, no answer has yet been provided that does not involve a UG stipulation.

On the *de se* theory of indexicality put forth in this paper, the person feature on a pronoun does not specify quantification over the members of the reference set at all.

Instead they specify (universal) quantification over the utterance's *speakers or addressees*. The person feature indicates a self-ascription condition of the form 'Every speaker/addressee self-ascribes ...':

(44) First and second person plurals on the De Se Theory.

- a. 1st person exclusive: 'Every speaker self-ascribes membership in the reference set.'
- b. 1st person inclusive: 'Every speaker and addressee self-ascribes membership in the reference set.'
- c. 2nd person: 'Every addressee self-ascribes membership in the reference set.'

Such a theory does not allow us, even in principle, to define a 'true second person plural', restricted to referring to addressees, or a 'true first person plural', restricted to referring to speakers. Such hypothetical pronouns are impossible in principle. A pronoun's grammatical specification does not involve quantification over the members of the reference set; but just such quantification is needed in order to designate that the reference set can include 'only addressees'. This explains the Associative Plural Generalization.

Assuming that plural number indicates a cardinality greater than one, and given the impossibility of a regular plural semantics, the *de se* theory actually allows for two possibilities for a language. The first is that first and second person plurals receive the associative interpretation described above, that is, the speaker or addressee self-ascribes membership in the (plural) reference set. The second is that a language could lack first

and second person plurals altogether. Presumably the former is favored by the clear utility of first and second person plurals for communication.

The existential statements in 43 are true, but they are epiphenomena not directly specified in the grammar. Rather, they follow from the following: (i) the self-ascription conditions in 33/34: the speakers (1st person exclusive), addressees (2nd person), or both (1st person inclusive), are self-ascribers, hence they refer via their self-notions; and (ii) the Self-Notion Axiom: for any agent *A*, *A*'s self-notion is a notion of *A*.

On the *de se* theory, the space of typological variation is determined by different constraints languages place on the speech act role of the self-ascriber, but not by grammatical constraints on the pronoun referent per se. The only constraint on the pronoun referent per se is the Self-Ascription Monopoly (the relevant part of which, as noted, may turn out to be superfluous under some theories of markedness). Being a principle, it operates uniformly in all languages, affecting the interpretation of all pronouns.

In Section 2.3 above we rejected the attempt to assimilate the semantics of person-marking to that of gender. A more apt comparison, given the *De Se* Theory, would be between person semantics and *gender-exclusive differentiation*— although the analogy is imperfect. In languages with gender-exclusive differentiation, male and female speakers have different speech forms, and are not normally allowed to speak the variety of the other gender. Differentiation of verb forms for male and female speakers of Koasati (Muskogean; spoken in Louisiana), for example, is governed by regular rules such as the following: if the female speaker's verb form ends in a nasalized vowel, the men's form has final [s] and no nasalization (Haas 1944, 143). Thus the verb form

meaning ‘he is saying’ is *kã* for a female speaker but *kas* for a male speaker; and the form meaning ‘I am not lifting it’ is *lakawtakkõ* for a female speaker but *lakawtakkos* for a male speaker (Haas 1944, 143). Note that the latter example happens to involve a first person subject. So an utterance of *lakawtakkõ* ‘I am lifting it’ allows one to infer that the subject denotation or ‘lifter’ argument is female, while *lakawtakkos* allows one to infer that the subject denotation or ‘lifter’ argument is male. But this inference about the biological gender of the subject argument is not directly represented in the grammar of Koasati gender differentiation described above. As just noted, these forms are grammatically differentiated according to the gender of the speaker, not for the gender of the grammatical subject’s denotation.^{xi}

Analogously, on the *de se* Theory proposed here, first and second person pronouns are not grammatically distinguished according to their denotation, but rather according to which speech act participant, the speaker or addressee, is designated as self-ascriber when the pronoun is used.

We have made two key claims: that personal pronoun indexicality is self-ascription; and that the value of the person feature (first / second / inclusive) indicates which speech act participant self-ascribes, instead of indicating which speech act participant the pronoun refers to, as the received view would have it. As long as we make those assumptions, then the Associative Plural Generalization follows as an immediate consequence. In essence, the explanation can be summed up as follows. The reason why there are no pronouns specialized for referring to ‘only addressees’ or ‘only speakers’ is because reference to ‘addressee’ and ‘speaker’ is not directly distinguished at all within pronoun systems.

6. Indexicality and Theory of Mind

6.1 Self-ascription and Theory of Mind

On the present account, the semantics of indexical pronouns breaks down into two components.

First, the person feature of the pronoun designates certain speech act participants as self-ascribers: [*spk*] designates the speakers as self-ascribers and [*addr*] designates the addressees as self-ascribers. Consider first person pronouns first, then second person. If Mary says, ‘I am beautiful’, she self-ascribes the property of beauty. We have seen one way of formalizing Mary’s self-ascribed belief as involving her self-notion. This is all the speaker needs to know about the semantics of first person pronouns, in order to understand the pronoun.

However, for the *addressee* to understand Mary’s use of the first person, he must infer Mary’s self-ascribed belief. He does so in roughly the same way that we as linguists have analyzed her utterance: he constructs a model of Mary’s belief state by applying the rules of the language. To construct a model of someone else’s belief state, an agent must exercise the human ability known as *theory of mind*, the cognitive ability to impute mental states to others and draw inferences from them (Premack and Woodruff, 1978).^{xii} Thus for an addressee to correctly interpret a first person utterance requires theory of mind, while the speaker can interpret (and therefore produce correctly) her own first person utterance without the need for theory of mind.

Similar reasoning applies to second person pronouns, only with the speech act participants reversed. If someone says to Mary, ‘You are beautiful’, and Mary believes

it, then once again Mary self-ascribes the property of beauty. This is all the addressee needs to know about the semantics of second person pronouns, in order to understand the pronoun. But for the *speaker* to understand a second person pronoun, and thus produce it in the appropriate contexts, requires that speaker construct a model of the addressee's belief state, which requires the speaker to exercise her theory of mind.

Let us summarize as follows:

(45) The two components of personal pronoun interpretation:

(i) *Self-ascription* is responsible for the correct interpretation of first person pronouns by the speaker and of second person pronouns by the addressee; and

(ii) The construction of a model of the other interlocutor's mental state by employing *theory of mind*, is responsible for the correct interpretation of first person pronouns by the addressee and second person pronouns by the speaker.

Table III indicates which combinations of speech act participant and personal pronoun type require theory of mind (ToM):

INSERT TABLE III ABOUT HERE

This two-component theory has important implications for language acquisition and developmental psychology. At a general level, populations with a lack or deficit of theory of mind are predicted to have special problems with first and second person pronouns. More specifically, such populations are expected to have particular problems

with the appropriate production (as speakers) of second person pronouns and the correct interpretation (as addressees) of first person pronouns. Two such populations with a theory of mind deficit are discussed below: typically developing young children (Section 6.2), and children with autism (Section 6.3).

6.2 Evidence from the acquisition of pronouns

The development of Theory of Mind in children has been addressed in hundreds of empirical studies (see Wellman, Cross, and Watson 2001 for a meta-analysis). Most studies employ the classic ‘false belief’ experimental paradigm (Wimmer and Perner 1983). A child is presented with a scenario such as the following:

Maxi puts his chocolate in the kitchen cupboard and leaves the room to play. While he is away (and cannot see) his mother moves the chocolate from the cupboard to a drawer. Maxi returns. Where will he look for his chocolate, in the drawer or in the cupboard? (Wellman, Cross, and Watson 2001, 655)

Crucially, Maxi’s belief diverges from reality and a correct answer by the subject depends on that false belief rather than the real situation. To answer correctly the child must have a notion of Maxi’s mental state, so a correct answer provides evidence of her mastery of Theory of Mind. Three-year-olds typically fail this test, while four- and five-year olds typically pass it. Children begin to answer correctly consistently around the age of 3.5 to 4 years of age (Sodian 2006, 97). As observed by Tager-Flusberg (2001, 174),

‘the dramatic change in performance on these kinds of tasks at about the age of 4 is one of the most robust findings in child development literature.’

While earlier research on Theory of Mind focused on the genesis of false-belief understanding, more recent work tends to view it as but one important milestone along a pathway in the development of Theory of Mind from infancy to adolescence (for an overview see Tager-Flusberg 2001, 178-9). Nonetheless, the important point for our purposes is that much of language development occurs *before* the child has mastered Theory of Mind. Children normally begin speaking long before age four. Single word utterances begin around the age of one, two-word combinations at about two years, and by the age of three a child is often producing fairly complex syntactic constructions such as subordinate clauses.

Given the two-component model of indexical pronouns in 45 above, we predict that during the period before theory of mind has (fully) developed, there should be relatively more correct use of pronouns by self-ascribers than by non-self-ascribers (see Table IV).

INSERT TABLE IV ABOUT HERE

In other words, we predict more instances of correct speaker-production of first person pronouns and correct addressee-comprehension of second person pronouns, since neither of these combinations require theory of mind. In contrast, during this same period we expect fewer instances of correct comprehension of first person pronouns by the addressee or production of second person by the speaker, since these combinations do

require theory of mind. Since the pronoun denoting the self-ascriber (whether speaker or addressee) is acquired before the pronoun denoting the non-self-ascriber, we call his prediction *Self-Ascribers First*:

- (46) *Self-Ascribers First*. A prediction regarding the order of pronoun acquisition: Correct pronoun use by Self-Ascribers precedes correct pronoun use by other speech act participants. Thus the order of acquisition is predicted to be: (i) first, production of first person pronouns and comprehension of second person pronouns; (ii) then comprehension of first person pronouns and production of second person pronouns.

Indeed, *Self-Ascribers First* is a robust finding from acquisition studies (Charney 1980; Chiat 1986). The following paragraph from Loveland (1984, 536) aptly summarizes the main findings.^{xiii}

It is generally agreed that *I/you* and their possessives are normally acquired in their non-anaphoric uses by age three. A typical pattern for the acquisition of these terms has now emerged (e.g. Bloom et al. 1975; Brown 1973; Deutsch and Pechmann 1978; Halliday 1975; Huxley 1970; Nelson 1975; Waryas 1973). *I/my/mine* are often produced early, apparently in reference to the self, though it is likely that many such productions are undifferentiated forms in which the pronoun is embedded in a fixed context. By contrast, *I/ my/mine* are not correctly comprehended at first (when used by others to refer to themselves), even if the

child can produce the terms. *You/your/yours* are rarely produced at all in the early stages of acquisition, though they are very often comprehended early (as applied to the child by others.) (Loveland 1984, 536)

In short, we find early *production* of first person and *comprehension* of second person, with other combinations developing later. One aspect of this pattern has appeared particularly mysterious. In language acquisition, comprehension normally precedes production, but for the first person forms the order is reversed: ‘As Charney points out, the production of *my* without comprehension seems illogical. The children would only be able to produce *my* in self-reference if they had already understood other speakers’ use of *my* as self-referring.’ (Chiat 1986, 347)

But as stated already, this pattern of acquisition is exactly what is expected if the ability to self-ascribe precedes the mastery of theory of mind. During the period before theory of mind is in place, successful pronoun use is favored for the self-ascribers, that is, for the speakers producing *I/my/mine* and the addressees comprehending *you/your/yours*.

In contrast, on the standard view it was hard to see why the acquisition of these pronouns should proceed in exactly this order. If first and second person pronouns were designated as referentially anchored to ‘speaker’ and ‘addressee’ respectively, then that anchoring would apply equally regardless of whether one were producing or comprehending the utterance.

In one revealing study, Loveland (1984) investigated the developmental relationship between spatial point of view and correct use of *I/you* pronouns by means of a cross-sectional and longitudinal study of 2-year-olds. Loveland concluded that ‘a

breakthrough in pronoun use comes at about the time the child learns that points of view can differ' (Loveland 1984:554), and specifically claimed that children master spatial point of view first, then apply it to pronoun comprehension and production.

Assuming the standard analysis of indexicals, it is unclear why the knowledge 'that points of view can differ' should be needed in order to understand indexical pronouns. On the standard theory of indexicals there is no reason why one would need to take the speaker's point of view in order to understand their utterance of a first person pronoun. All that one needs to know is (i) who is speaking, and (ii) that the first person pronoun refers to whoever is speaking. But on the present two-component *de se* theory of indexicals, a first person pronoun is not specified for referring *to* the person who is speaking. Instead, it is specialized for self-ascription *by* the speaker. To interpret the pronoun, the addressee builds a model of the speaker's belief state. Hence the difficulty with pronouns observed in young children follows directly from their weakness at precisely that model-building ability, also known as theory of mind.

6.3 Evidence from autism

Childhood autism is a severe developmental disorder characterized by specific social, emotional, cognitive, and linguistic impairments. While many children with autism are mentally retarded, others have IQ's within the normal range. The most influential theory, the *Theory of Mind Hypothesis of Autism*, attributes autism to a deficit or lack of a Theory of Mind (Baron-Cohen, Leslie, and Frith 1985; Tager-Flusberg 2001; Tager-Flusberg and Joseph 2005). In a seminal study, Baron-Cohen et al (1985) carried out puppet play experiments using Wimmer and Permer's false belief paradigm (see

Section 5.1 above) on three sets of children: a high functioning subgroup of children with autism (mean age 11;11), a group of much younger normal children (mean age 4;5), and a group of children with Down syndrome (mean age 10;11). Among the three groups, only the children with autism failed to impute mental states to others, even though their mental age (IQ) was greater than that of the two control groups. Baron-Cohen et al (1985) concluded that a lack of theory of mind was independent of retardation and specific to autism. These results have been replicated, although many autism researchers now favor a more nuanced view in which theory of mind is not always lacking entirely among children with autism, but rather it is impaired to varying degrees in different children; and moreover this impairment may be but one important component of autism, along with such other factors as a deficit in executive functions (for discussion see Tager-Flusberg, 2001; Tager-Flusberg and Joseph, 2005).

The Theory of Mind Hypothesis of Autism has been supported by non-linguistic evidence such as the Baron-Cohen et al (1985) experiment described above. More specifically, Hobson and Hobson (2007) argue that children with autism have impairments in the ‘self-other mapping’ needed for what is called *identification* with another person, that is, roughly the ability to see the world through another’s eyes. Along with the pronoun reversals, children with autism show ‘imitative “reversal errors,” for example, when they copy a model who displays her hands facing outward, by facing their own hands inward.’ (Jessica A. Hobson and R. Peter Hobson 2007, 414). Recent research on signing errors by deaf autistics supports those observations (Shield and Meier 2010).

Turning to linguistic symptoms of autism, perhaps the most prominent feature of the speech of children with autism, noted throughout the literature on autism beginning with the original description of the syndrome (Kanner 1943), is a special difficulty with the use of first and second person pronouns, ‘to a degree that seems out of keeping with other aspects of their language development’ (Lee, R. Peter Hobson, and Chiat 1994, 156).

The problem specifically involves errors of person and not, for example, case-marking errors: children with autism have a tendency to reverse the person feature of pronouns, referring to themselves as *you* and to their addressee as *I*. In one study, for example, all first and second person pronouns were extracted from transcripts of the speech of autistic and Down syndrome children, and analyzed for person reversal errors (*you* for *I*, etc.) and case errors (*me* for *I*, etc.) (Tager-Flusberg 1994, 184). Over 13% of the autistic children’s pronouns were person reversal errors (220 out of 1,673 pronouns), while the Down syndrome children made no person reversal errors (0 out of 2,270 pronouns). Meanwhile, case errors broke in the opposite direction: Down syndrome children made case errors on 1.23% of pronouns (28 out of 2,270) while children with autism made case errors on only 0.12% of pronouns (2 out of 1,673). Clearly children with autism have a special problem with the correct use of first and second person pronouns. The question is why.

The pronoun reversal errors have been cited as evidence that children with autism are confused about speaker/listener roles, a confusion related to impaired social cognitive functioning (Tager-Flusberg, 1994):

Autistic children have great difficulty understanding that different people have distinct conceptual perspectives— that people perceive, interpret, remember, and respond to situations in unique ways. This kind of conceptual perspective-taking ability is required in order to understand the different roles of speaker and listener (Loveland 1984), and this is reflected in the way pronouns are used in ongoing discourse. (Tager-Flusberg, 1994)

In short, deficits in the Theory of Mind make it difficult for children with autism to use first and second person indexical pronouns correctly. Much as we noted at the end of the previous section with regard to (typically developing) two-year-olds, it is not clear why this ‘perspective-taking ability’ should be needed in order to understand pronouns, assuming the standard theory. On the standard theory of indexicals there is no need to take the speaker’s perspective in order to understand their utterance of a first person pronoun. All that one needs to know is (i) who is speaking, and (ii) that a first person pronoun refers to whoever is speaking. But on the present two-component theory of indexicals, a first person pronoun is not specified for referring *to* the person who is speaking. Instead, it is used for self-ascription *by* the speaker. To interpret the pronoun, the addressee builds a model of the speaker’s belief state. Hence the difficulty with pronouns observed in children with autism follows directly from their deficit of theory of mind.

7. Conclusion

At its core the present proposal is very simple: that first and second person pronouns are grammatically specialized for self-ascription by the speaker and addressee, respectively. The idea that first person pronouns are self-ascriptive is an old one, and in that sense the present analysis belongs to a family of previous solutions to the problem of the essential indexical. But those earlier accounts assumed that first person pronouns, in addition to being specialized for self-ascription *by* the speaker, were also grammatically restricted for reference *to* the speaker. I have proposed to rid the grammar of that further condition on pronoun reference. The seemingly untenable consequence of simplifying the semantics of indexical pronouns in this way— namely that the grammar provides only the speaker, and no one else, with the means to understand a first person pronoun— turns out not to pose a problem. Other people understand the speaker's use of the first person pronoun by a combination of grammatical and extragrammatical means: they build a model of the speaker's mental state, and easily infer, based on the grammatical rule of self-ascription, that a first person pronoun refers to the speaker who produces it. All of the above also applies, *mutatis mutandis*, to second person pronouns and addressees.

Rather than posing an intractable problem, this simplification of the semantics of person is the linchpin connecting the observations from studies of typology, acquisition, and childhood autism. It explains why first and second person plural pronouns lack the regular plural semantics found with vast majority of other plural nominals: if first and second person pronouns are not grammatically specified for reference to speaker and hearer, then the plural forms of such pronouns can hardly be grammatically restricted to

such reference, as would be required for regular plural semantics. The special difficulties with indexical pronouns experienced by young children, and especially by children with autism, follows from a deficit of the theory of mind ability in those populations. This convergent evidence from multiple different domains constitutes strong support for the proposal.

References

- Anand, Pranav. 2006. *De de se*. PhD dissertation, Cambridge MA: MIT.
- Anand, Pranav, and Andrew Nevins. 2004. Shifty operators in changing contexts. *Proceedings of SALT XIV*, ed. by R. Young, 20-37. Ithaca, NY: CLC Publications, Cornell University.
- Asher, Nicholas. 1986. Belief in discourse representation theory. *Journal of Philosophical Logic* 15: 127-189.
- Baron-Cohen, Simon, Alan M. Leslie, and Uta Frith. 1985. Does the autistic child have a “theory of mind”? *Cognition* 21: 37–46.
- Bloom, Lois, Patsy Lightbown, Lois Hood, M. Bowerman, M. Maratsos, and M. P. Maratsos. 1975. Structure and variation in child language. *Monographs of the Society for Research in Child Development* 40: 1-97.
- Bobaljik, Jonathan David. 2008. Missing persons: A case study in morphological universals. *The Linguistic Review* 25: 203-230.
- Brown, Roger. 1973. *A first language: the early stages*. Cambridge, MA: Harvard University Press.
- Castañeda, Hector-Neri. 1977. Perception, belief, and the structure of physical objects and consciousness. *Synthese* 35: 285-351.
- Charney, Rosalind. 1980. Speech roles and the development of personal pronouns. *Journal of Child Language* 7: 509-528.
- Chiat, Shulamuth. 1986. Personal pronouns. *Language acquisition*, ed. by Paul Fletcher and Michael Garman, 339-355. Cambridge: Cambridge University Press.

- Chierchia, Gennaro, and Sally McConnell-Ginet. 2000. *Meaning and grammar: an introduction to semantics*. Cambridge, MA: MIT Press.
- Comrie, Bernard. 1980. Review of Greenberg, Joseph H. (ed.), *Universals of Human Language*, Volume 3: Word Structure (1978). *Language* 56: 834-838.
- Cooper, Robin, and Jonathan Ginzburg. 1996. A compositional situation semantics for attitude reports. *Logic, Language and Computation*, ed. by Jerry Seligmann and Dag Westerståhl, 1:151-166. Vol. 1. Stanford, CA: CSLI Publications.
- Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.
- . 2000. *Number*. Cambridge: Cambridge University Press.
- . 2005. Suppletion in personal pronouns: Theory versus practice, and the place of reproducibility in typology. *Linguistic Typology* 9: 1-23.
- Corbett, Greville, and Marianne Mithun. 1996. Associative forms in a typology of number systems: evidence from Yup'ik. *Journal of Linguistics*: 1-17.
- Crimmins, Mark. 1992. *Talk about beliefs*. Cambridge, MA: MIT Press.
- Crimmins, Mark, and John Perry. 1989. The prince and the phone booth: reporting puzzling beliefs. *The Journal of Philosophy*: 685-711.
- Cysouw, Michael. 2003. *The paradigmatic structure of person marking*. Oxford: Oxford University Press.
- Dalrymple, Mary, and Ronald M. Kaplan. 2000. Feature indeterminacy and feature resolution. *Language* 76: 759-798.
- Deutsch, Werner, and Thomas Pechmann. 1978. Ihr, dir, or mir? On the acquisition of pronouns in German children. *Cognition* 6: 155-168.
- van Ditmarsch, Hans, Wiebe van der Hoek, and Barteld Kooi. 2006. *Dynamic epistemic*

- logic*. Berlin: Springer.
- Fauconnier, Gilles. 1985. *Mental spaces*. Cambridge, MA: MIT Press.
- Forchheimer, Paul. 1953. *The category of person in language*. Berlin: Walter de Gruyter.
- Geach, Peter T. 1972. On beliefs about oneself. *Logic matters*, 128-129. Berkeley: University of California Press.
- Givón, Talmy. 1970. The resolution of gender conflicts in Bantu conjunction: when syntax and semantics clash. *Papers from the Sixth Regional Meeting, Chicago Linguistic Society, April 16-18, 1970*, 250-261. Chicago: Chicago Linguistic Society.
- Greenberg, Joseph H. 1988. The first person dual as an ambiguous category. *Studies in Language* 12: 1-18.
- Grice, H. P. 1975. Logic and conversation. *Syntax and Semantics: Volume 3*, ed. by Peter Cole and Jerry L. Morgan, 41-58. New York: Academic Press.
- Haas, Mary R. 1944. Men's and women's speech in Koasati. *Language* 20: 142-149.
- Halliday, M. A. K. 1975. *Learning how to mean: Explorations in the development of language*. London: Edward Arnold.
- Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. PhD dissertation, University of Massachusetts at Amherst.
- Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Malden, MA: Blackwell.
- Hobson, Jessica A., and R. Peter Hobson. 2007. Identification: The missing link between joint attention and imitation? *Development and Psychopathology* 19: 411-431.
- Huxley, R. 1970. The development of the correct use of subject personal pronouns in two

- children. *Advances in psycholinguistics*, ed. by G.B. Flores d'Arcais and W. J. M. Levelt, 141-165. Amsterdam: North-Holland.
- Jackendoff, Ray. 1990. *Semantic structures*. MIT Press: Cambridge, MA.
- Kamp, Hans. 1971. Formal properties of 'now'. *Theoria* 37: 237-273.
- . 1981. A theory of truth and semantic representation. *Formal methods in the study of language*, ed. by Jeroen A. G. Groenendijk, T. M. V. Janssen, and Martin B. J. Stokhof, 277-322. Amsterdam: Mathematical Center Tract 135.
- Kanner, Leo. 1943. Autistic disturbances of affective contact. *Nervous Child* 2: 217-250.
- Kaplan, David. 1977. Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. *Themes from Kaplan*, ed. by Joseph Almog, John Perry, and Howard Wettstein, 481-566. NY and Oxford: Oxford University Press.
- Kratzer, Angelika. 2009. Building a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40: 187-237.
- Krifka, Manfred. 2001. Quantifying into question acts. *Natural Language Semantics* 9: 1-40.
- Kripke, Saul. 1979. A puzzle about belief. *Meaning and Use*, ed. by Avishai Margalit, 239-283. Dordrecht: Reidel Publishing Company.
- Landman, Fred. 2000. *Events and plurality: The Jerusalem lectures*. Dordrecht, NL: Kluwer Academic Publishers.
- Lee, Anthony, R. Peter Hobson, and Shulamuth Chiat. 1994. I, you, me, and autism: An experimental study. *Journal of Autism and Developmental Disorders* 24: 155-176.
- Lewis, David. 1979. Attitudes de dicto and de se. *The Philosophical Review* 88: 513-543.

- Link, Godehard. 1998. *Algebraic semantics in language and philosophy*. Stanford, CA: CSLI Publications.
- Loveland, Katherine A. 1984. Learning about points of view: spatial perspective and the acquisition of 'I/you'. *Journal of Child Language* 11: 535-556.
- Milne, A. A. 1926. *The complete tales of Winnie-the-Pooh*. New York: Dutton Children's Books.
- Mithun, Marianne. 1988. Lexical categories and the evolution of number marking. *Theoretical Morphology: Approaches in Modern Linguistics*, ed. by Michael Hammond and Michael Noonan, 211-234. San Diego, CA: Academic Press.
- Moravcsik, Edith. 2003. A semantic analysis of associative plurals. *Studies in Language* 27: 469-503.
- Moravcsik, Edith A. 1978. Agreement. *Universals of Human Language*, ed. by Joseph H Greenberg, 331-74. Stanford: Stanford University Press.
- Nelson, Katherine. 1975. The nominal shift in semantic-syntactic development. *Cognitive Psychology* 7: 461-479.
- Noyer, Rolf. 1992. *Features, positions and affixes in autonomous morphological structure*. PhD dissertation, Cambridge, MA: MIT.
- Nunberg, Geoffrey. 1993. Indexicality and deixis. *Linguistics and Philosophy* 16: 1-43.
- O'Grady, William, Michael Dobrovolsky, and Mark Aronoff. 1993. *Contemporary Linguistics: an Introduction*. New York: St. Martin's.
- Perry, John. 1979. The problem of the essential indexical. *Noûs* 13: 3-21.
- Plank, Frans. 1985. Die Ordnung der Personen. *Folia linguistica* 19: 111-176.
- Premack, David, and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind?

- The Behavioral and Brain Sciences* 1: 515-526.
- Richard, Mark. 1983. Direct reference and ascriptions of belief. *Journal of Philosophical Logic* 12: 425-452.
- Scha, Remko. 1981. Distributive, collective and cumulative quantification. *Formal methods in the study of language*, ed. by J. A. G. Groenendijk, T. M. V. Janssen, and M. J. B. Stokhof. Amsterdam: Mathematisch Centrum.
- Schlenker, Philippe. 2003a. Indexicality, logophoricity, and plural pronouns. *Research in Afroasiatic Grammar II: Selected Papers from the Fifth Conference on Afroasiatic Languages, Paris, 2000*, ed. by Jacqueline Lecarme, 409–428. Amsterdam: John Benjamins.
- . 2003b. A plea for monsters. *Linguistics and Philosophy* 26: 29-120.
- Searle, John R. 1983. *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Shield, Aaron, and Richard Meier. 2010. Visual perspective taking in sign language: Evidence from deaf children with autism. *Paper presented at the annual meeting of the Linguistic Society of America*. Baltimore, MD.
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. *Features and projections*, ed. by Pieter Muysken and Henk C. van Riemsdijk, 163-232. Dordrecht: Foris.
- Simon, Horst J. 2005. Only you? Philological investigations into the alleged inclusive-exclusive distinction in the second person plural. *Clusivity*, ed. by Elena Filimonova, 113-150. Amsterdam: John Benjamins.

- Smith-Stark, T. Cedric. 1974. The plurality split. *Chicago Linguistic Society*, 10:657-672. Vol. 10.
- Sodian, Beate. 2006. Theory of Mind— The case for conceptual development. *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind*, ed. by Wolfgang Schneider, Ruth Schumann-Hengsteler, and Beate Sodian, 95-130. Mahwah, NJ: Routledge.
- Sokolovskaja, N. K. 1980. Nekotorye semantičeskie universalii v sisteme ličnyx mestoimenij. *Teorija i tipologija mestoimenij*, ed. by I. F. Vardul', 84–102. Moscow: Nauka.
- Tager-Flusberg, Helen. 1994. Dissociations in form and function in the acquisition of language by autistic children. *Constraints on language acquisition: studies of atypical children*, ed. by Helen Tager-Flusberg, 175-194. Hillsdale, NJ: Lawrence Erlbaum Associates.
- . 2001. A reexamination of the theory of mind hypothesis of autism. *the development of autism— perspectives from theory and research*, ed. by Jacob A Burack, Tony Charman, Nurit Yirmiya, and Philip R Zelazo, 173-193. Mahwah, NJ: Routledge.
- Tager-Flusberg, Helen, and Robert M Joseph. 2005. Theory of mind, language, and executive functions in autism: a longitudinal perspective. *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind*, ed. by Wolfgang Schneider, Ruth Schumann-Hengsteler, and Beate Sodian, 239-257. Mahwah, NJ: Routledge.
- Waryas, C. L. 1973. Psycholinguistic research in language intervention programming: the

- pronoun system. *Journal of Psycholinguistic Research* 2: 221-237.
- Wechsler, Stephen. 2008. Elsewhere in gender resolution. *The nature of the word—essays in honor of Paul Kiparsky*, ed. by Kristin Hanson and Sharon Inkelas, 567-586. Cambridge: MIT Press.
- Wellman, Henry M, David Cross, and Julianne Watson. 2001. Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development* 72: 655–684.
- Wimmer, H., and J. Perner. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13: 103–128.

Notes

* Earlier versions of this paper were presented at the *Cognitive Science Program Lecture Series*, University of Texas (August 2008), the *Symposium on Verb Agreement in Spoken and Signed Languages* at the *LSA Annual Meeting* (January 2009), and the *5th Workshop on Discourse Structure*, University of Texas (November 2009). I wish to thank Elizabeth Coppock, Greg Carlson, David Beaver, Richard Meier, John Beavers, and Colin Bannard for their comments on earlier drafts. Richard Meier led me to see the relevance of theory of mind. Elizabeth Coppock suggested the connection to autism, helped with the formalization, and was a source of insight into these issues generally. Greg Carlson gave me very useful comments on two earlier drafts. I would also like to thank Hans Kamp and Mark Sainsbury, for discussion of *de se* reference and indexicality. And as always, thank-you, Marie and Jonas, for showing me what you and I mean to each other.

ⁱ Standard English second person pronouns (*you, your*) do not distinguish number, although reflexives do (*yourself ~ yourselves*) and there are various dialectal plurals such as *y'all, you guys*, and *yous*. In this paper we indicate an English second person pronoun's number value by using one of the forms that distinguish number as just mentioned, or by explicitly stating the number interpretation.

ⁱⁱ This discussion closely follows Corbett (1991, 279ff).

ⁱⁱⁱ The actual analysis of person features proposed above does not involve these rules. They are presented here to show that the semantics of person features does not allow the same range of variation as the semantics of gender features.

^{iv} The astute reader will recognize this story as a variant of ‘Pooh and Piglet Go Hunting and Nearly Catch a Woozle’, by A. A. Milne (1926, 32-41).

^v Nunberg (1993, 20) notes that indexicals sometimes behave like descriptions. He gives the example of a condemned prisoner saying, *I am traditionally allowed to order whatever I like for my last meal*. This is interpreted roughly as ‘The condemned prisoner is traditionally allowed to order whatever he likes for his last meal.’ Notice that the utterance is still a self-ascription: the speaker must self-ascribe the presupposition of *instantiating* (to use Nunberg’s term) the ‘condemned prisoner’ property. It is that self-ascriptive part of the interpretation that is our focus here.

^{vi} Doxastic puzzles similar to the sugar-spilling story, but involving the second person pronouns, are discussed by Richard (1983).

^{vii} Note that we restrict our attention to *uses* of the pronouns, as opposed to *mentions* of them, the latter found in direct quotation, for example; and we focus on *indexical* pronouns, as opposed to special anaphorically bound uses of first and second person such as ‘fake indexicals’ (Kratzer 2009) and ‘shifted interpretations’ of local pronouns in languages like Amharic, Slave, and Zazaki (Schlenker 2003b; Schlenker 2003a; Anand and Nevins 2004; Anand 2006).

^{viii} Crimmins and Perry (1989) calls this function *Of*.

^{ix} Crimmins (1992) posits a basic relation *Believes*(A, t, p, τ) which holds between an agent A , a time t , a proposition p , and a belief structure τ . I have suppressed the time

variable for simplicity, and switched the order of the arguments for the belief structure and its content.

^x The *character* of an expression is defined as a function from contexts to intensions (David Kaplan 1977).

^{xi} Gender-exclusive differentiation for the addressee as well as the speaker has been observed in at least one language, namely Biloxi (Siouan; extinct) (O'Grady, Dobrovolsky, and Aronoff 1993, 433).

^{xii} Such an ability is considered a 'theory' because the mental states are not directly observable and the system can be used to make predictions (Premack and Woodruff 1978:515).

^{xiii} As noted in the quote from Loveland (1984), it has been claimed that many of the early productions of the first person are 'undifferentiated forms in which the pronoun is embedded in a fixed context' (see Chiat 1986). If so then those cases do not involve distinct first person pronominal morphemes, but rather are effectively portmanteau morphs or suppletive forms. The resolution of that issue affects the morphological analysis but has no effect on the argument presented in this paper, which concerns only the first person *interpretation*. As long as the child has acquired one or more forms that express self-reference, it makes no difference whether the pronominal features are encoded by an analyzable morpheme or not.

Possible		Attested
1+2	speaker(s) and addressee(s); no others	'inclusive'
1+2+3	speaker(s), addressee(s), and other(s)	
1	speaker(s) only	'exclusive'
1+3	speaker(s) and other(s); addressee(s) excluded	
2	addressee(s) only	'second person'
2+3	addressee(s) and other(s)	
3	other(s) only	'third person'

Table I.

The seven logically possible meta-persons, and the four attested pronoun types

Possible		Attested	Binary features
1+2	speaker(s) and addressee(s); no “others”	‘inclusive’	[+spk, +hr]
1+2+3	speaker(s), addressee(s), and other(s)		
1	speaker(s) only	‘exclusive’	[+spk, –hr]
1+3	speaker(s) and other(s); addressee(s) excluded		
2	addressee(s) only	‘second person’	[–spk, +hr]
2+3	addressee(s) and other(s)		
3	other(s) only	‘third person’	[–spk, –hr]

Table II.

A ‘UG Solution’ to the Associative Plural Generalization

	Production by speaker	Comprehension by addressee
1st pers. pronouns	<i>self-ascriber</i>	<i>ToM</i>
2nd pers. pronouns	<i>ToM</i>	<i>self-ascriber</i>

Table III.

Need for Theory of Mind (ToM) by interlocutors using indexical pronouns.

	Production by speaker	Comprehension by addressee
1st pers. pronouns	<i>self-ascriber</i> EARLY	<i>ToM</i> LATE
2nd pers. pronouns	<i>ToM</i> LATE	<i>self-ascriber</i> EARLY

Table IV.

Predicted order of mastery of indexical pronoun production and comprehension.