HISTORY AND PHILOSOPHY OF LOGIC, 20 (2000), 251–265

Quantification in English is Inherently Sortal

EDWARD L. KEENAN
Department of Linguistics, University of California, Los Angeles, Hilgard Avenue, Los Angeles CA 90024, USA; e-mail: ekeenan@ucla.edu

Received 6 May 1999

Within Linguistics the semantic analysis of natural languages (English, Swahili, for example) has drawn extensively on semantical concepts first formulated and studied within classical logic, principally first order logic. Nowhere has this contribution been more substantive than in the domain of quantification and variable binding. As studies of these notions in natural language have developed they have taken on a life of their own, resulting in refinements and generalizations of the classical quantifiers as well as the discovery of new types of quantification which exceed the expressive capacity of the classical quantifiers. We refer the reader to Keenan and Westerståhl (1997) for an overview of results in this area. Here, we focus on one property of quantification in natural language—its inherently sortal nature—which distinguishes it from quantification in classical logic.

1. From logic to linguistic analysis

Within Linguistics a primary goal of semantics is to formulate a compositional semantic interpretation for the expressions of a given natural language. This, of course, presupposes a grammar which defines the set of expressions pretheoretically judged grammatical by native speakers. At the time of writing, grammars for English that have been proposed are incomplete—they fail to generate some expressions which speakers judge grammatical, and unsound—they generate some expressions not judged grammatical by native speakers. Nonetheless, our understanding of the syntax of certain simple fragments of English is clear enough that it makes sense to ask for a compositional semantics for those fragments. We focus here on issues concerning the semantic analysis of quantificational structures in natural language.

Classical Logic (CL) provides semantic representations like (1b) and (2b) for the English sentences (Ss) in (1a) and (2a) respectively:

a. All poets daydream b. $\forall x(\text{Poet}(x) \rightarrow \text{Daydream}(x))$ (1)

a. Some poets daydream b. $\exists x(\text{Poet}(x)\ \&\ \text{Daydream}(x))$ (2)

Ignoring properties associated with tense (present, past) and aspect (generic, perfective), this semantic analysis is correct in the sense that systematic use of such representations correctly captures certain judgements of semantic relatedness given by native speakers. As an example, an unambiguous English sentence P is understood to entail an unambiguous sentence Q *iff* their semantic representations P′ and Q′ are such that Q′ is interpreted as True in all models in which P′ is.

Moreover this semantic analysis enables linguists to represent a variety of semantic distinctions which are difficult to understand and represent in the absence of a systematic representation of quantification. The following are three cases: first, English Ss like (3) are semantically ambiguous; they can be understood in two logically distinct ways, as represented in (3a) and (3b), using some obvious abbreviations.
Each student in the class read some play over the vacation

a. $\forall x(Sx \rightarrow \exists y(Py\ \&\ xRy))$ b. $\exists y(Py\ \&\ \forall x(Sx \rightarrow xRy))$ (3)

Of course (3a) can be true if no two students read a play in common, as long as for each student there is some play that that student read. In (3b) by contrast, there must be at least one play which is such that each student read that play. Thus, using representations which differ with respect to the relative scope of the quantifiers, $\forall x$ and $\exists y$ permits a natural representation of the semantic ambiguity in (3).

Second, in early work in generative grammar verb phrase (VP) coordination was derived from coordinate Ss by, roughly, eliminating repeated occurrences of Noun Phrases (NPs). It was understood that the derived Ss had the same meaning as those they were derived from, thus satisfying Compositionality. As an example, (4b) was to be derived from (4a), with which it is logically synonymous.

a. John laughed and John cried   b. John laughed and cried   (4)

Obviously such a Conjunction Reduction transformation fails to be paraphrastic when the NPs are quantified and an appropriate coordinate conjunction is selected. Thus, (5a) and (5b) are not logical paraphrases, as is clear from their CL representations:

a. Some student laughed and some student cried

$(\exists x(Sx \ \& \ Lx) \ \& \ \exists x(Sx \ \& \ Cx))$

b. Some student laughed and cried $\exists x(Sx \ \& \ Lx \ \& \ Cx)$ (5)

Of course, if *and* is replaced by *or* in (5a, b) the resulting Ss are logically equivalent. Analogous claims hold for, *Either every student came early or every student left late* and, *Every student either came early or left late*. The two Ss are not paraphrases, though the results of replacing *or* by *and* are. Again, these semantic facts are clearly accounted for using CL representations.

Thus, CL helps us to see that the naive linguistic analysis of expressions like (5b) is problematic for Compositionality, as the semantic interpretation of the derived expression does not stand in a regular semantic relation to the one it is derived from. Sometimes it is logically equivalent to it and sometimes not. To get the correct truth conditions for the non-paraphrastic cases we need to see not only the identity of the conjunction used (*and* vs *or*) but also the choice of NP (*every student* vs *some student*).

Third, in a similar vein, early work in generative grammar sometimes purported to derive expressions by replacing full, non-pronominal (NPs) by pronouns when the full NP was identical to another appropriate NP occurrence but, semantic problems comparable to those for Conjunction Reduction above arose, (6a, b) are not paraphrases as is clear from their CL representations.

a. All poets admire all poets $\forall x \forall y((Px \ \& \ Py) \rightarrow xAy)$

b. All poets admire themselves $\forall x(Px \rightarrow xAx)$ (6)

So again, if a grammar of English generates (6b) by replacing the second occurrence of *all poets* in (6a) by the reflexive pronoun *themselves* we find that Compositionality is hard to satisfy. Merely knowing the models that satisfy (6a) is not sufficient to identify those that satisfy (6b), as the latter is a proper subset of the former.

In all these ways the representations of Classical Logic have proven insightful in the semantic analysis of natural language expressions. It might then seem surprising that the, often informally presented, semantic representations used by linguists for quantificational expressions in natural language differ from those of CL.

1.1 *Some linguistic objections to the CL analysis*

All approaches to English syntax agree that in (1) the sequence *all poets* forms a syntactic constituent. It consists of the Determiner (Det) *all* and the (plural marked) noun *poets*. The VP *daydream* forms the other constituent of (1). We expect by Compositionality then, that the semantic interpretation of the entire S is given in terms of the interpretation of *all poets* and that of *daydream* and, thus, that these constituents have a semantic interpretation. However, in (1b), the CL translation of (1a), there is no syntactic constituent which represents the meaning of the NP *all poets*. Rather, the noun *poet* is ripped away from its Det *all* and is treated as a one place predicate.

Moreover, tied to the linguist's respect for syntactic constituency here is the intuition that the semantic roles of the noun *poet* and that of the VP *daydream* are quite different. We can think of both as denoting properties that individuals may or may not have. Yet, the noun property serves to limit the range of objects we are talking about, specifically those we are quantifying over, whereas the VP presents the property we are predicating of those objects (in accordance with the constraints determined by the Det *all*). By contrast, in (1b) the variable x is understood to range over all the individuals in the universe of discourse. We may fairly read it in rough English as, 'For all individuals x, if x is a poet then x daydreams'.

So, (1a) and (1b) differ in that in (1a) we are just talking about poets, whereas in (1b) we are talking about everything, though what we predicate of those objects is now expressed by a boolean compound of formulas built from the original noun and the original VP. It is something of an embarrassment to this intuitive difference in meaning that, modulo tense and aspect, (1b) does adequately represent the truth conditions and entailment relations of (1a). Perhaps this is an accident of the example. (2a, b) suggest this may be the case. (2b), like (1b), quantifies over all objects in the universe of the model but, it incorporates the noun into its predicate differently, by using *and* rather than *if-then*. Will yet different Determiners require yet further boolean connectives in combining the noun and the VP? Are there enough boolean connectives to accommodate the variety of English Dets? We see below that the answer is negative and, thus, that natural languages, in distinction to standard first order languages, are *inherently sortal*. However, we anticipate. Let us first consider the direct interpretation of NPs of the form Det + Noun.

2. From linguistics to logic

Traditionally, we think of subject-predicate Ss such as *John daydreams* as ones in which the predicate *daydreams* is the general term and the subject *John* the specific one. This is captured extensionally by treating a possible predicate denotation as a set of possible subject denotations and, we represent the truth in a model of *John daydreams* by saying that the object *John* denotes is an element of the set of objects *daydreams* denotes.

Yet, as Frege realized, this general-specific distinction is cut the other way when we consider quantified NP subjects such as *all poets*, *some poets*, *no poets*, etc., rather than simple proper names. Now it is the subject phrase which denotes the general term and the predicate the more specific one. That is, extensionally, the set of possible quantified NP denotations corresponds to sets of one place predicate denotations.

To see the idea behind this claim we take a simple example and show how to construct 2^n extensionally distinct NP denotations, where n is the number of extensionally distinct VP denotations. In fact, we can take the NPs to be proper nouns and just consider their logically distinct boolean compounds in *and*, *or*, *not*, and *neither ... nor ...* Consider, for example, a universe with just three elements, a, b, c denoted say by Adam, Bill, and Chris. Now, adjusting number marking on the verb appropriately, consider the eight Ss that result when X in (7a) is replaced by one of the eight NPs in (7b):

a. X daydreams

b. 1 Adam and Bill and Chris
2 Adam and Bill but not Chris
3 Adam and Chris but not Bill
4 Adam but neither Bill nor Chris
5 Bill and Chris but not Adam
6 Bill and neither Chris nor Adam
7 Chris and Adam but not Bill
8 Neither Adam nor Bill nor Chris (7)

For X = (b. 1) we compute that (7a) is true iff *daydreams* denotes {a, b, c}. When X is (b. 2) it is true iff *daydreams* denotes {a, b}, and so on to (b. 8), where (7a) is true iff *daydreams* denotes the empty set. In this way then we see that the eight NPs in (7b) are logically distinct, each one corresponding a single possible VP denotation. Now, take any subset of the NPs in (7b) and form their disjunction: e.g. either Adam and Bill but not Chris, or both Bill and Chris but not Adam, or neither Adam nor Bill or Chris. Clearly when X is such a disjunction (7a) is true iff *daydreams* denotes one of the sets denoted by one of the disjuncts. So disjunctions of distinct subsets of these NPs determine logically distinct NPs, so the number of logically distinct NPs corresponds to the number of sets of extensionally distinct VP denotations. In the case at hand we build 2^8 logically distinct NPs. (Note we are really just constructing NPs in disjunctive normal form, in analogy to the way this is done in propositional logic; see Keenan and Faltz 1985).[1]

Of course in forming logically distinct NPs we can have recourse to ones that are not boolean compounds of proper nouns. Consider the NP like *every student and no non-student*. Setting X to be this NP, (7a) above is true iff the objects who daydream are exactly the students. So this NP can denote any of the eight possible denotations given by (b. 1)–(b. 8) above according to the set *student* denotes. Moreover, interpreting *student* as {a, b, c} in the example above we can again form eight logically distinct NPs using quantifiers and exception phrases, as in *every student*, *every student but Adam*, *every student except Adam and Chris*, ..., *no student but Chris*, ..., *no student*.

Now to say that NPs determine sets of VP denotations says that we can treat NPs semantically as functions mapping VP denotations into {True, False}. Call such functions *generalized quantifiers*. Consider, for example, *all poets*. Semantically it maps a set B, which we sometimes call the *predicate set*, to True iff each object in the set of poets is in B. That is, writing denotations in upper case, (ALL POET) (B) = True iff POET $\subseteq$ B. More generally, for A, B any sets, (ALL A) (B) = True iff A $\subseteq$ B, and

1 The analogy is exact. Possible NP denotations will shortly be taken to be generalized quantifiers, which constitute a complete, atomic boolean algebra. The set of possible proper noun denotations, the individuals, is a set of complete, free generators for this set, just as the set of denotations of so called atomic formulas in sentential logic is a set of free generators for boolean algebra of logical equivalence classes of formulas.

this in turn says that we can interpret *all* as a function ALL which maps a set A to the generalized quantifier ALL(A). In this way we give a compositional interpretation to (1a) as in (8):

All poets daydream

ALL POET DAYDREAM

ALL(POET)

ALL(POET)(DAYDREAM) (8)

Note that this compositional interpretation dispenses with variable binding and does not introduce the extraneous connective 'if-then'. This remains true when all is replaced by any of the other Dets whose denotations are given transparently in (9).

a. (ALL BUT ONE)(A)(B) = True iff $|A \cap B| = 1$
b. SOME(A)(B) = True iff $A \cap B \neq \varnothing$
c. NO(A)(B) = True iff $A \cap B = \varnothing$
d. (MORE THAN TEN)(A)(B) = True iff $|A \cap B| > 10$
e. (THE TEN)(A)(B) = True iff $|A| = 10$ and $A \subseteq B$
f. MOST(A)(B) = True iff $2.|A \cap B| > |A|$
g. (MORE THAN TWO OUT OF THREE)(A)(B) = True iff $3.|A \cap B| > 2.|A|$ (9)

These results are linguistically very satisfying: Ss which differ syntactically just by a lexical item (*all* for *some*, etc.) differ semantically just by the denotations of those lexical items. So, the difference in interpretation between *All poets daydream* and *Some poets daydream* is obtained by replacing ALL by SOME in (8). In addition, directly interpreting NPs as generalized quantifiers eliminates the problem of introducing different boolean connectives for different Dets—*if-then* for *all* and, *and* for *some*.

However, the linguistic advantages of interpreting NPs as generalized quantifiers run much deeper than uniformity and simplicity of interpretation. We now have a format in which to present and study denotations of natural language Determiners. We can study what properties they have in common, we can discern linguistically natural classes and, we can formulate and test whether English Dets are sortally reducible.

2.1. *Some semantic classes of English Dets*

For simplicity of presentation, we assume we are given an arbitrarily chosen non-empty universe E of objects held fixed throughout the discussion unless stated otherwise. GQ(E), the set of generalized quantifiers over E, is the set of functions from P(E), the set of subsets of E, into {True, False}; and the Dets under discussion denote functions from P(E) into GQ(E). Functions from P(E) into GQ(E) will be called *possible Determiner denotations.* We claim later that not all of these are actual; there are some denotation constraints that all English Dets satisfy.

Let us see first how the distinction between universal and existential quantifiers shows up in our generalized quantifier format. While we no longer translate *all* and *some* in such as way as to introduce distinct boolean connectives, the semantic different that those connectives represented still exists as a condition on the functions which universal and existential Dets satisfy.

2.1.1. *Generalized existential Dets in English.* The existential Det, *some* in English is *intersective* in the sense that whether *Some As are Bs* is True is decided just by checking $A \cap B$, the set of As that are Bs. We do not have to know anything about As that are not Bs or Bs that are not As. We just check that the set of As that are Bs is non-empty. If so the S is true; if not it is false. Equally, NO is intersective: whether NO(A)(B) = True is decided just by checking whether $A \cap B$ is empty.

Definition. A possible Det denotation D is *intersective* iff for all subsets A, A′, B, B′ of E,

$$\text{if } A \cap B = A' \cap B' \text{ then } D(A)(B) = D(A')(B')$$

So, an intersective D cannot distinguish among arguments which have the same intersection. Here are two groups of intersective Dets in English (an intersective Det being one whose denotation in every model is an intersective function as per Definition 1):

(i) some, no, a/an, not a, not a single, hardly any, practically no, almost no, a dozen, more than ten, fewer than ten, exactly/at least/nearly/approximately ten, a few, several, between five and ten, not more than ten, at least ten and not more than twenty, either fewer than ten or else more than a hundred, just finitely many, infinitely many

(ii) no ... but John, more male than female, at least two male (10)

The Dets in (10i) are not merely intersective they are *cardinal* in the sense that whether a function D they denote maps a pair A, B of sets to True just depends on the cardinality of $A \cap B$. D does not have to know what the elements of $A \cap B$ are, it merely checks how many elements it has, e.g. *fewer than ten* is cardinal since (FEWER THAN 10)(A)(B) = True iff $|A \cap B| < 10$. Formally,

Definition. A possible Det denotation D is *cardinal* iff for all subsets A, A′, B, B′ of E,

$$\text{if } |A \cap B| = |A' \cap B'| \text{ then } D(A)(B) = D(A')(B')$$

Cardinal Dets are studied in Keenan and Moss (1985). Here, we note two points used later: first, boolean compounds of cardinal (intersective) Dets are themselves cardinal (intersective). For example, *not more than ten* is cardinal since *more than ten* is; *at least two and not more than ten* is cardinal since each conjunct is. In general, boolean compounds in *and*, *or*, and *not* of Dets, of whatever sort, not just intersective ones, are given *pointwise* as follows. Where we write $\vee$ for the interpretation of *or*, and $\neg$ for that of *not*:

a. $(F \wedge G)(A)(B) = F(A)(B) \wedge G(A)(B)$

b. $(F \vee G)(A)(B) = F(A)(B) \wedge G(A)(B)$

c. $(\neg F)(A)(B) = \neg(F(A)(B))$ (11)

The objects on the right of the = sign in (11) are truth values, and the $\wedge$, $\vee$, and $\neg$ operations are the usual truth functions. So, from (11a), we see that (12a, b) are logically equivalent:

a. Most but not all students read the Times

b. Most students read the Times but it is not the case that all students read the Times (12)

And secondly, the cardinal Dets include the two constant functions: **T**, which maps all A, B to True, and **F**, which maps all A, B to False. Note that these functions are denotable:

a. At least zero = **T** b. Fewer than zero = **F** (13)

(13a) holds since (AT LEAST ZERO)(A)(B) = True for all sets A, B; and (13b) holds since (FEWER THAN ZERO)(A)(B) = False, all A, B. One checks directly that **T** and **F** are both intersective, in fact both cardinal.

The expressions in (10ii) have a different character from those in (10i).[2] One might doubt whether they should be considered Determiners at all. However, before rejecting them out of hand let us see just what is intended. Here are some Ss illustrating their uses:

a. No student but John jogs during lunch
b. More male than female students play football
c. At least two male and not more than five female students won prizes. (14)

(14a) says in effect that the students who came early consist just of John. Treating *no ... but John* as a discontinuous Det we obtain the correct truth conditions using,

Definition 3. (NO ... BUT JOHN)(A)(B) = True iff $A \cap B = \{John\}$.

Clearly *no ... but John* is intersective—it yields the same value at pairs A, B and A′, B′ which have the same intersection. Yet it is not cardinal. If $A \cap B = \{John\}$ and $A' \cap B' = \{Bill\}$ then the two intersections have the same cardinality but (NO ... BUT JOHN) is true in the first case and false in the second.

So, if we treat *no ... but John* as a Det it is intersective but, not cardinal. Should we treat it as a Det? There are in fact some linguistic reasons for doing so. Suppose, for example, that we thought of *but John* in *no student but John* as forming a constituent with *student* to the exclusion of the Det *no*. Then, *student but John* would be a syntactic

2 The expressions in (10i) lack both the 'constant' and the 'logical' properties of logical constants and, classical quantifiers have both. A succinct way to capture the essential idea is to note that the possible denotations of these expressions fail to respect permutations of the underlying universe. This is a notion that can be used to characterize the 'logical' elements of any type. For the Det case at hand: let π be a permutation of the universe E. Extend π to a function π^* from P(E) to P(E) by setting:

$$\pi^*(A) = \{\pi(a) | a \in A\}.$$

Observe, omitting the straightforward proofs, that (1) π^* is a bijection of P(E), whence for all $A \subseteq E$, $|\pi^*(A)| = |A|$, and (2) π^* commutes with the boolean operations on P(E). That is, $\pi^*(A \cap B) = \pi^*(A) \wedge \pi^*(B)$ and $\pi^*(\neg A) = \neg(\pi^*(A))$, where of course $\wedge$ and $\neg$ on the right hand side of these equations refers to the relevant operations in the truth value algebra. Then,

Definition A Possible Det denotation D over a universe E is *permutation invariant* (PI) iff for all permutations π of E, all subsets A, B of E, $D(\pi^*(A))(\pi^*(B)) = D(A)(B)$.

Then one shows, by example, that the Dets in (10ii) may denote D that fail to be PI. Moreover, being PI + intersective characterizes the property of being cardinal (over finite universes).

Theorem: For E finite, a possible Det denotation D is cardinal iff D is intersective and PI.

$\Rightarrow$ That D is intersective is immediate from the definition of cardinal. Let π be a permutation of E. We must show that for A, B arbitrary, $D(\pi^*(A))(\pi^*(B)) = D(A)(B)$. But since $|A \cap B| = |\pi^*(A \cap B)| = |\pi^*(A) \cap \pi^*(B)|$ the result follows since D is cardinal. Note that this direction does require that E be finite.

$\Leftarrow$ Let D be PI and intersective, with E finite. Suppose $|A \cap B)| = |A' \cap B'|$. We must show that $D(A)(B) = D(A')(B')$. Since D is intersective and $X \cap Y = E \cap (X \cap Y)$ for all X, $Y \subseteq E$, we have that $D(X)(Y) = D(E)(X \cap Y)$, all X, $Y \subseteq E$. And since E is finite, $|\neg(A \cap B)| = |\neg(A' \cap B')|$. Let π_1 be a bijection: $A \cap B \sim A' \cap B'$ and let π_2 be a bijection: $\neg(A \cap B) \sim \neg(A'B')$. Then $\pi = \pi_1 \cup \pi_2$ is a bijection of E with $\pi^*(A \cap B) = A' \cap B'$. Thus $D(A)(B) = D(E)(A \cap B) = D\pi^*(E)\pi^*(A \cap B) = D(E)(A' \cap B') = D(A')(B')$, as was to be shown.

unit of the sort that Dets would combine with to form full NPs. However, this yields massively incorrect predictions, as most choices of Det are ungrammatical here (as indicated by *):

*two students but John, *most students but John, *the ten students but John (15)

Essentially, only *no* and *every* are grammatical here. Thus, the prenominal Det and the exception phrase *but John* do not occur independently, which is predicted if we treat them as forming a syntactic unit into which the noun *student* is infixed. We then favour treating *no ... but John* and *every ... but John* as (discontinuous) Determiners.

In the case of (14b), *more male than female* (and infinitely many variants thereof: *many more male than female*, *ten more male than female*, *twice as many male as female*, *fewer male than female*, *exactly as many male as female* ...) we treat adjectives like *male* and *female* as absolute functions from sets (common noun extensions) to sets, as follows:

Definition 4. A function F from P(E) to P(E) is *absolute* iff for all $A \subseteq E$,

$$F(A) = A \cap F(E)$$

So to say that *male* is absolute is to say that the male artists are the artists who are male individuals, which is correct. And, we interpret *more male than female* by:

$$\text{(MORE MALE THAN FEMALE)(A)(B)} = \text{True iff } |\text{MALE(A)} \cap B| > |\text{FEMALE(A)} \cap B| \quad (16)$$

So, *More male than female students play ball* is True iff the number of male students who play ball is greater than the number of female students who play ball. Observe that this Det is intersective. If $A \cap B = A' \cap B'$ then the two sets whose cardinality we compare on the right in (16) are the same using A, B throughout or using A′, B′ throughout, replacing A, B with A′, B′ respectively preserves cardinality, so the inequality holds in one case iff it holds in the other. Observe, for F absolute and $A \cap B = A' \cap B'$, that:

$$\begin{aligned} F(A) \cap B &= (A \cap F(E)) \cap B && \text{F is absolute} \\ &= (A \cap B) \cap F(E) && \text{Associativity \& Commutativity of } \cap \\ &= (A' \cap B') \cap F(E) && \text{Assumption } A \cap B = A' \cap B' \\ &= (A' \cap F(E)) \cap B' && \text{Associativity \& Commutativity of } \cap \\ &= F(A') \cap B' && \text{F is absolute} \end{aligned} \quad (17)$$

However, *more male than female* may denote a function which fails to be cardinal. With John male and Mary female set, A = B = {John} and A′ = B′ = {Mary}. Then, $|A \cap B| = |A' \cap B'|$ but (MORE MALE THAN FEMALE)(A)(B) = True and (MORE MALE THAN FEMALE)(A′)(B′) = False. So MORE MALE THAN FEMALE IS NOT cardinal. Similar arguments show that TWO MALE in (14c) is intersective but not cardinal.

There is then a *prima facie* case that English presents syntactically complex Dets which are intersective but not cardinal. And in any case intersectivity is a property of many English Dets, both simplex and complex. Observe now the following Proposition which is the reflection at the level of Generalized Quantifiers of the introduction of *and* in the classical translation of the existential quantifier. It also leads to the result that intersective Dets are sortally reducible (a notion we define shortly).

Proposition 1. For D a possible Det denotation over a universe E,

$$\text{D is intersective iff for all A,B} \subseteq \text{E, D(A)(B)} = \text{D(E)(A} \cap \text{B)}$$

proof:

$\Rightarrow$ Clearly $A \cap B = E \cap (A \cap B)$ so $D(A)(B) = D(E)(A \cap B)$ by the intersectivity of D.

$\Leftarrow$ Let X, X', Y, Y' be arbitrary subsets of E with $X \cap Y = X' \cap Y'$. Show $D(X)(Y) = D(X')(Y')$. Now:

$$\begin{aligned} D(X)(Y) &= D(E)(X \cap Y) && \Leftarrow \\ &= D(E)(X' \cap Y') && \text{assumption} \\ &= D(X')(Y') && \Leftarrow \end{aligned}$$

Proposition 1 guarantees the logical equivalence of (18a, b) below, give that *more than ten* is intersective. Moreover, *more than ten* can be replaced by any intersective Det, including 'unexpected' ones like *exactly as many male as female*, preserving logical equivalence (though singular and plural marking may have to be adjusted).

a. More than ten students are talking
b. More than ten individuals are students and are talking (18)

Now Proposition 1 tells us that when D is intersective, the use of the noun argument A to restrict the set of objects quantified over is not essential in the sense that we can replace A by E, thus quantifying over all elements of the universe and, compensate for the original restriction by incorporating the noun property into the predicate in some boolean way. For intersective Dets the compensation is simply by intersection. Let us now formulate the notion of *sortal reducibility* and see that intersective Dets have this property.

Definition 5. Let D be a possible Det denotation over a universe E. We say that D is *sortally reducible* iff there is a two place boolean function h satisfying:

$$\text{for all A,B} \subseteq \text{E, D(A)(B)} = \text{D(E)h(A,B)}$$

Clearly all intersective Dets are sortally reducible: just choose h to be intersection. Thus, in Ss of the form [[Det N] VP] with Det intersective, we see that restricting the domain of quantification to the set denoted by the N is not an essential restriction. We can replace the N denotation by the entire universe, that is we can quantify over everything and, compensate by building a new predicate property as a boolean function of the original N denotation and the original predicate property (denoted by the VP).

We turn now to the generalized universal quantifiers in English. We show that they are also sortally reducible. Then we show that given a certain, very general constraint on natural language Determiner denotations, the only sortally reducible Dets in English are the generalized existential and the generalized universal ones. For the many other cases which we show exist, to see that the restriction of the domain of quantification to the set denoted by the noun argument of Det is essential; it cannot be paraphrased away by quantifying over all individuals and compensating in some boolean way by enriching the original predicate with that determined by the original noun argument.

2.1.2. *Generalized Universal Dets in English.* Our development here parallels that of the Generalized Existential Dets in English. First, recall that we have interpreted English *all* by that possible Det denotation ALL given by: ALL(A)(B) = True iff $A \subseteq B$. An equivalent statement, which makes the parallel with intersective Dets more apparent, is:

$$\text{ALL(A)(B)} = \text{True iff } A-B = \varnothing \tag{19}$$

(Clearly A is a subset of B if removing all the Bs from the As leaves nothing, and conversely). Now (19) makes it clear that the value ALL assigns to a pair A, B of sets is decided by a property of A–B. We define:

Definition 6. A possible Det denotation D is *co-intersective* iff for all subsets A, A′, B, B′ of E,

$$\text{if } A-B = A'-B' \text{ then } D(A)(B) = D(A')(B')$$

And we shall take co-intersectivity as the defining property of the generalized universal Dets, just as we took intersectivity as the defining property of the generalized existential Dets. Clearly ALL is co-intersective. So are the denotations of the following:

a. every, each, nearly all, all but ten, all but at most ten, all but finitely many
b. every ... but John, almost every ... but John, every ... except John and Bill, (20)

Denotations for the a-group above are easy to state (modulo vagueness, and treating *every* and *each* as synonyms of *all*). Here are some examples, which show that they are co-intersective:

a. (ALL BUT TEN)(A)(B) = True if $|A-B| = 10$
b. (ALL BUT AT MOST TEN)(A)(B) = True iff $|A-B| \leqslant 10$
c. (ALL BUT FINITELY MANY)(A)(B) = True iff A–B is finite (21)

(Note: we might think of the universal quantifier *all* as *all but zero*). We observe that the Dets in the a-group are not only co-intersective, they are *co-cardinal* in the sense that the value they assign to a pair A, B of sets is decided just by checking the cardinality of A–B. We leave the definition of *co-cardinal* to the reader. And we observe that the expressions in the b-group are co-intersective (but not co-cardinal), as in:

a. Every student but John plays football
b. (EVERY ... BUT JOHN)(S)(P) = True iff S–P = {John}. (22)

And clearly (EVERY ... BUT JOHN) is co-intersective, as whether it maps a pair S, P to True is decided just by looking at S–P. However, since it must see more than just the number of elements in S–P, it must know what they are, it is not co-cardinal. We note in passing that the trivial Det denotations **T** and **F** are co-intersective, in fact co-cardinal (as well as intersective and cardinal). In fact they are the only functions that are both intersective and co-intersective.

Observe now that the co-intersective Dets are reducible but, not by *and* (intersection), as was the case for the generalized existential Dets, but by *if-then*, which we write in the booleanly more familiar form $\neg A \bigcup B$ rather than $A \rightarrow B$.

Proposition 2. A possible Det denotation D is co-intersective iff for sets A, B

$$D(A)(B) = D(E)(\neg A \bigcup B)$$

proof:

$\Rightarrow$. $E-(\neg A \bigcup B) = E \bigcap (\neg\neg A \bigcap \neg B) = A-B$, whence by the co-intersectivity of D, $D(E)(\neg A \bigcup B) = D(A)(B)$

$\Leftarrow$. Let D satisfy the equation above for all A, B. We show that D is co-intersective. Let A, A′, B, B′ such that $A-B = A'-B'$. Then $D(A)(B) = D(E)(\neg A \bigcup B) = D(E)\neg\neg(\neg A \bigcup B)) = D(E)\neg(A \bigcap \neg B) = D(E)\neg(A-B) = D(E)\neg(A'-B') = \ldots = D(A')(B')$, the missing steps being those used in the previous steps, in reverse, replacing A by A′, B by B′.

Corollary 3: Prop 2 entails immediately that co-intersective Dets are reducible via the function h which maps each $(A, B) = (\neg A \bigcup B)$.

2.2. *Non-classical quantifiers in English*

We have taken the properties of intersectivity and co-intersectivity as the basis for identifying classes of English Dets which have the existential and universal quantifiers as special cases. Note that even if we limit ourselves to the cardinal and co-cardinal elements of these classes we still go beyond the expressive power of first order logic. As an example, simple compactness arguments show that the intersective *just finitely many* and the co-intersective *all but finitely many* are not first order definable[3].

English however presents a great many Determiner expressions which are neither intersective nor co-intersective. Here are three types, of which the last is the most convincing. First, non-trivial boolean compounds of intersective with co-intersective Dets typically form complex Dets which are neither intersective (int) nor co-intersective (co-int). For example, *some but not all* (*As are Bs*) is not int, since it requires knowledge of A–B to check that not all As are Bs. It is not co-int since it requires knowledge of $A \bigcap B$ to verify that some As are Bs.

Second, *presuppositional* Dets like *both*, *neither*, *the ten*, *the ten or more*, *John's ten* (*or more*) given below are neither int. nor co-ints.

a. $\text{BOTH}(A)(B) = \text{True iff } |A| = 2 \text{ and } A \subseteq B$

b. $\text{NEITHER}(A)(B) = \text{True iff } |A| = 2 \text{ and } A \bigcap B = \varnothing$

c. $(\text{THE TEN})(A)(B) = \text{True iff } |A| = 10 \text{ and } A \subseteq B$

d. $(\text{JOHN's TEN})(A) = (\text{THE TEN})(A \bigcap \{x \in E | \text{JOHN HAS } x\})$ (23)

Clearly *the ten* is not int, since if we just know which As are Bs we cannot tell how many As there are. Nor can we if we just know which As are not Bs, so *the ten* is not co-int.

Third, and highly productive in English, are the *proportional* Dets. They look at a pair A, B of sets and make claims about the proportion of As that are Bs. Here are two fairly simple examples (interpreting *most* in the sense of *more than half* and *seven out of ten* in the sense of *at least seven out of ten*):

a. $\text{MOST}(A)(B) = \text{True iff } 2.|A \bigcap B| > |A|$

b. $(\text{SEVEN OUT OF TEN})(A)(B) = \text{True iff } 10.|A \bigcap B| \geqslant 7.|A|$ (24)

3 For example, assuming that each positive integer n has a name in English, let K be the set of Ss of the form, 'There are at least n cats on the mat', for each positive integer n, together with the S, 'There are just finitely many cats on the mat'. Clearly each finite subset of K has a model and, equally clearly K itself has no model. Hence compactness fails, so any language including these Ss interpreted in the intended way fails to be first order.

most fails to be int since if all we know is which As are Bs, and hence how many As are Bs, we still don't know whether that number comes to more than half the number of As. Similarly MOST is not co-int since merely knowing which, and so how many, As are not Bs does not suffice to tell us the As that are Bs constitute more than half the As.[4] We define:

Definition 7. A possible Det denotation D is *proportional*[5] iff for all A, A′ B, B′ $\subseteq$ E,

$$\text{if } |A \cap B|/|A| = |A' \cap B'|/|A'| \text{ then } D(A)(B) = D(A')(B')$$

Here are some examples of proportional Dets in English. They include mundane fractional and percentage expressions,

most, more/less than half the, exactly/almost half the, at least a third of the, between one third and two thirds of the, a majority of the
at least/at most/exactly/less than ten per cent of the, between ten and twenty per cent of the, about/nearly ten per cent of the
at least/more than/exactly/almost/about seven out of ten (25)

We note that with only a few exceptions proportionality Dets are not (co-)int.[6] The reason is that given the noun set A and the predicate set B evaluating D(A)(B) requires knowledge of both A and $A \cap B$ (from which A–B is computable as $A-(A \cap B)$). Yet English Dets do not seem to require more knowledge than this. The statement that for each universe E, a possible Det denotation need know at most which objects are As and which of those are Bs, is known as Conservativity:

Definition 8. A possible Det denotation D is *conservative* iff for all A, B, B′ $\subseteq$ E,

$$\text{if } A \cap B = A \cap B' \text{ then } D(A)(B) = D(A)(B').$$

Thus, when D is conservative then for any A the generalized quantifier D(A) can't see the difference between predicate properties B and B′ that have the same intersection

4 In attempting to sortally reduce *most* speakers I've consulted tend to try to assimilate it to the co-intersective class, rendering *Most As are Bs* as *For most x, if x is an A then x is a B.* Yet this is clearly incorrect. In a model with 100 individuals, 10 of whom are students and just three of whom are vegetarians, the S, *Most students are vegetarians* is clearly false, as at most only three of the ten are. But the sentence *For most x, if x is a student then x is a vegetarian* is clearly true, as it holds for sure (vacuously) for 90 of the 100 individuals in the universe.

5 It is preferable (though it clouds the underlying intuition) to avoid the use of division in the statement since we want to consider the case where A or A′ is $\varnothing$. So just write $n \cdot m' > n' \cdot m$ instead of $n/m > n'/m'$.

6 The non-trivial exceptions include (and perhaps are limited to) the traditional square of opposition: *no*, expressible as *less than* $l/|E|$, and its complement *some*, both intersective; and *all*, and its complement, *not all*. We do not have an exact count of the non-trivial proportional Det denotations over a given (even finite) universe. Yet most Det denotations are not (co)-intersective. The map sending each intersective D to D(E) is provably an isomorphism from INT(E), the set of intersective functions over E, to GQ(E), any E. Since GQ(E) is the set of functions from P(E) into {True,False} its cardinality is 2 raised to the power $2^{|E|}$. Similarly CO-INT(E) is isomorphic to GQ(E) by the map sending each co-intersective D to D(E). As only **T** and **F** are both intersective and co-intersective $|\text{INT(E)} \cup \text{CO-INT(E)}| = 2$ raised to the power $2^{|E|+1}$, less the two elements that were counted twice. However, Keenan and Stavi (1986) show not only that the total number of conservative Det denotations (see later) is 2 raised to the power $3^{|E|}$, they show that for E finite, each of these functions is denotable by some English Determiner (usually syntactically complex). As an example, in a universe with just three elements there are just 510 possible Det denotations that are either intersective or co-intersective. There are 2^{27} or over 60 million that are conservative. So most possible Det denotations lie outside the (co)-intersective classes.

with A. And we claim that all natural language Dets are conservative. To test whether a Det is conservative use Proposition 4 (usually taken as the definition of Conservativity).

Proposition 4. A possible Det denotation D is conservative iff

$$\text{for all } A, B \subseteq E,\ D(A)(B) = D(A)(A \cap B)$$

We leave the proof to the reader. Using Proposition 4 one checks that an arbitrary Det *blik* is conservative by checking that *Blik As are Bs* is true in the same conditions as *Blik As are As that are Bs.* So the conservativity of properly proportional Dets such as *seven out of ten* is illustrated by observing the logical equivalence of *Seven out of ten students are vegetarians* and *Seven out of ten students are students who are vegetarians.* Even a tortured expression such as *more of John's than of Bill's* passes the conservativity test, as Ss like (26a) are clearly true in the same conditions as (26b).

a. More of John's than of Bill's cats are black

b. More of John's than of Bill's cats are cats that are black (26)

Conservativity holds since the predicates of the two Ss differ just in that one repeats information already contained in the noun property and, so doesn't add anything new. Indeed, presented as in Proposition 4, Conservativity may seem trivial. Are there possible Det denotations that fail to have this property? The answer is a resounding 'Yes!'. Here is one example:

Let E have at least two elements a, b; let D be that possible Det denotation given by

$$D(A)(B) = \text{True iff } |A| = |B|. \tag{27}$$

Then D is not conservative. $D(\{a\})(\{b\}) = \text{True}$ but $D(\{a\})(\{a\} \cap \{b\}) = D(\{a\})(\varnothing) =$ False. More generally (see Keenan and Stavi 1986) one computes that for any E, the total number of possible Det denotations is 2 raised to the power $4^{|E|}$, whereas the number of those which are conservative is 2 raised to the power $3^{|E|}$. So, in a model with just two elements there will be $2^{16} = 65{,}536$ possible Det denotations, only $2^9 = 512$ of which are conservative.

Despite the strength of Conservativity however, most conservative functions, even over a finite universe, are not definable in first order logic:

Proposition 5. Dets of the form *more than n/m*, for $1 \leqslant n < m \leqslant |E|$ are not first order definable even over finite universes E.

Barwise and Cooper (1981) sketch the proof for *more than 1/2*; the techniques used in Westerståhl (1989) enable one to handle the more general case (and many others) in *Proposition 5.*

We are now in a position to show:

Theorem 6: Given a universe E, a conservative Det denotation D is sortally reducible iff D is intersective or D is co-intersective.

Proof. A succinct but not very user friendly proof can be found in Keenan (1993). Here is a longer but more helpful one. We have already shown the right to left direction of the theorem. So let D be conservative and sortally reducible. We show that D is int or co-int. Now to say that D is sortally reducible is to say that for some two place boolean function h, $D(A)(B) = D(E)(h(A, B))$, all $A, B \subseteq E$. There are just 16 such functions, so we may give a proof by cases. Here, in set notation, are eight of these functions.

$$\begin{array}{ll} h_1(A)(B) = E; & h_5(A)(B) = A-B \\ h_2(A)(B) = A; & h_6(A)(B) = \neg A \cap B \\ h_3(A)(B) = B; & h_7(A)(B) = \neg A \cap \neg B \\ h_4(A)(B) = A \cap B & h_8(A)(B) = (A-B) \cup (B-A) \end{array} \tag{28}$$

The other 8 are, in effect, the complements of these. Formally, for $1 \leqslant i \leqslant 8$, set $g_i(A)(B) = \neg(h(A)(B))$. For example, $g_1(A)(B) = \neg E = \varnothing$.

Case 1. Suppose that D reduces via h_1. That is, for all A, B $D(A)(B) = D(E)(h(A, B)) = D(E)(E)$. This says that D is constant. That is, $D = \mathbf{T}$ or $D = \mathbf{F}$, according as $D(E)(E) =$ True or $D(E)(E) =$ False. And in each case D is int (also co-int). Similarly if D reduces via g_1. Then $D(A)(B) = D(E)g_1(A, B) = D(E)\varnothing)$, so again D is constant and thus int.

Case 2. Let D reduce via h_2. Then $D(A)(B) = D(E)(h_2(A, B)) = D(E)(A) = D(E)(h_2(E, A)) = D(E)(E)$; so again D is constant and thus int. If D reduces via g_2 then $D(A)(B) = D(E)(g_2(A, B)) = D(E)(\neg A) = D(E)(g_2(E, \neg A)) = D(E)(\varnothing)$, so D is constant and thus int.

Case 3. Let D reduce via h_3. Then $D(A)(B) = D(A)(A \cap B)$, by conservativity, $= D(E)(h_3(A, A \cap B)) = D(E)(A \cap B)$, whence D is int: if $A \cap B = A' \cap B'$ then $D(A')(B') = D(E)(A' \cap B') = D(E)(A \cap B)$. Similarly, if D reduces via g_3 then $D(A)(B) = D(E)(g_3(A, A \cap B)) = D(E)\neg(A \cap B)$, so again D is int.

Case 4. If D reduces via h_4 or g_4 then D is clearly intersective.

Case 5. If D reduces via h_5 or g_5 then d is clearly co-intersective.

Case 6. Let D reduce via h_6. So $D(A)(B) = D(E)(h_6(A, B)) = D(E)(\neg A \cap B) = D(E)(h_6(\neg E \cap (\neg A \cap B))) = D(E)(\varnothing)$, so D is constant and thus int. Similarly, one shows that if D reduces via g_6 then $D(A)(B) = D(E)(E)$ and so again D is constant and thus int.

Case 7. Similar to case 6. Let D reduce via h_7. Then $D(A)(B) = D(E)(h_7(A, B)) = D(E)(\neg A \cap \neg B) = D(E)(h_7(E, \neg A \cap \neg B)) = D(E)(\neg E \cap (\neg E \cap \neg B)) = D(E)(\varnothing)$, so D is constant, and so int. Similarly, if D reduces via g_7 then $D(A)(B) = D(E)(E)$ and so is constant.

Case 8. Let D reduce via h_8. Then $D(A)(B) = D(A)(A \cap B)$, by conservativity, $= D(E)(h_8(A, A \cap B)) = D(E)(A-(A \cap B) \cup (A \cap B)-A) \quad D(E)(A-B) \cup \varnothing) = D(E)(A-B)$, whence D is co-intersective. And, finally, if D reduces via g_8 then $D(A)(B) = D(A)(A \cap B) = D(E)(g_8(A, A \cap B) = D(E)\neg h_8(A, A \cap B) = D(E)(\neg(A-B))$, whence again D is co-intersective.

This exhausts the cases proving the theorem. Note that the conservativity of D was used only in cases 3 and 8.

3. Conclusion

We have shown here that quantification in English is inherently sortal, in the sense that for many quantifiers Q, sentences of the form in (29a) are not logically equivalent to any of the form in (29b), where the dots indicate some boolean compound and x ranges over the entire universe of discourse:

$$\text{a. } [[Q+N]+P] \qquad \text{b. } Qx(\ldots Nx \ldots Px \ldots) \tag{29}$$

Specifically, we have seen that logical equivalence obtains just when Q is either intersective or co-intersective. For properly proportional quantifiers, such as *most*, *exactly half the*, *two out of three*,... the use of the Noun to restrict the domain of quantification to the set denoted by the N is essential. There is no logical paraphrase of the form in (29b).

We end with a caveat: whether a quantifier is sortally reducible and whether it is first order definable are independent properties, though it happens that the properly proportional quantifiers fail both conditions: they are neither sortally reducible nor are they first order definable, not even over finite universes. But the quantifiers *just finitely many* and *all but finitely many* are sortally reducible but not first order definable. And the quantifiers *both*, *neither*, and *the ten* are first order definable but not sortally reducible.

References

Barwise, J. and Cooper R. 1981. 'Generalized Quantifiers and Natural Language', *Linguistics and Philosophy*, **4**, 159–219.

Keenan, E. L. 1993. 'Natural Language, Sortal Reducibility and Generalized Quantifiers', *Journal of Symbolic Logic*, **58**, 314–325.

Keenan, E. L. and Faltz, L. 1985. *Boolean Semantics for Natural Language*, Reidel, Dordecht.

Keenan, E. L. & Moss, L. 1985. 'Generalized Quantifiers and the Expressive Power of Natural Language', in J. van Benthem and A. ter Meulen (eds), *Generalized Quantifiers*, Foris, Dordrecht, pp. 73–124.

Keenan, E. L. and Stavi, J. 1986. 'A Semantic Characterization of Natural Language Determiners', *Linguistics and Philosophy*, 9, 253–326.

Keenan, E. L. and Westerståhl, D. 1997. Generalized Quantifiers in Linguistics and Logic, in J. van Bentham and A. ter Meulen (eds), *Handbook of Logic and Language*, Elsevier: Amsterdam, pp. 837–893.

Westerståhl, D. 1989. Quantifiers in Formal and Natural Languages, in D. Gabbay and F. Guenthner (eds), *Handbook of Philosophical Logic, Volume IV*, Reidel, Dordrecht, pp. 1–131.