LING 564, Formal Semantics 01/18/05
Useful Concepts from Set Theory, Part 1

I. Sets and How to Describe Them

A set is a collection of objects, entities, individuals, etc – anything we can sensibly think of as a piece of the universe can be part of a set. The group of things which comprise a set can be highly concrete, as in (1), highly abstract (2), a mixture of both (3), or things whose existence is somewhat obscure (4)[1]:

(1) a set containing all and only: Andy's watch, the chair in the NW corner of this room, and Andy's station wagon

(2) a set containing all and only: the numbers 3, 4, 5, and pi

(3) a set containing all and only the integers larger than 5, and Andy's station wagon.

(4) Pocohontas, Shakespeare, Merlin, and Batman.

Each "item in a set" is termed a <u>member</u> of that set. There are two standard ways to specify what's in a particular set. The members can simply be listed, if there is a finite number of them. This is done in (1), (2), and (4), crudely, and the more proper notation is to enclose the descriptions/names of the members within curly brackets {}. So, the right representation for (1,2,4) is actually:

(1') {Andy's watch, the chair in the NW corner of this room, and Andy's station wagon}

(2') {3, 4, 5, pi}

(4') {Pocohontas, Shakespeare, Merlin, Batman}

The second way to specify a set is via abstraction, which is giving a general description of a property (or properties) that hold(s) of all and only the members of the set. Since the set in (3) is infinite in size, we can't just list them all. But we can specify the set as follows:

---

[1] There's a fudge here, since the existential status of each of these is different. Pocohontas was a real woman, although her characterization in folk tales is somewhat innaccurate; Shakespeare was a real person, but there's some doubt as to which guy wrote the plays, so there's an indeterminacy of what the name "Shakespeare" actually refers to; Merlin may or may not have existed; and Batman exists, but only in a superhero universe. We'll return to puzzles posed by referential terms later on; for the moment, let's be relaxed and allow them into our sets.

(2') $\{x \mid x$ is Andy's station wagon or x is an integer $> 5\}$

When specifying a set via a description of the property/properties true of its members, we talk about an indefinitely large number of things. So, we need to use variables to stand in for the items in question. We'll return to (2'), but for a second let's deal with a different set:

(5) $\{x \mid x$ is an integer $> 5\}$

Here, the variable is x, and you should read (5) as "the set of all x's, such that each x is an integer greater than five". The description in (5) does not tell you how many (if any) objects actually satisfy the property described – it just tells you what property something would have to have to be in that set.

However, any set does have a certain number of members. This is so even if, for some sets, the number is an order of infinity, like (5), or unknown, as in (6), or unknowable, as in (7):

(6) $\{y \mid y$ is a blade of grass on Andy's back lawn at 3:45 PM Tuesday, January 18th, 2005 CE$\}$

(7) $\{z \mid z$ is a person who will be born on planet Earth sometime between January 1, 2100 and Dec 31st, 5100$\}$

Regardless of how a set is specified, the relationship between its members and the set is constant: for a given item, it's either a member of that set, or it isn't. No halfway cases are allowed.

If object a is a member of set S, we write

$a \in S$

If it isn't, we write

$a \notin S$

II. Cardinality

The size of a set – the number of members it has – is the cardinality of the set. This is a number. The notation for "the cardinality of set X" is:

(8) $|X|$

To apply the concept to some of the earlier examples, consider (2') and (4'). What are their cardinalities?

(2') {3, 4, 5, pi}

(4') {Pocohontas, Shakespeare, Merlin, Batman}

(9) $|\{3, 4, 5, \text{pi}\}| = 4$

(10) $|\{\text{Pocohontas, Shakespeare, Merlin, Batman}\}| = 4$

Sets can be given names. By convention, upper case Roman letters are used to name sets. So, we could call the set specified in (2') A, and the set named in (4') B.

(11) $A = \{3, 4, 5, \text{pi}\}$

(12) $B = \{\text{Pocohontas, Shakespeare, Merlin, Batman}\}$

Since

(13) $|A| = 4$

and

(14) $|B| = 4$

then

(15) $|A| = |B| = 4$

That is, the <u>size</u> of these two sets is the same – they each have four members. However … it's totally invalid to think

(16) $A = B$

(16) talks about the sets themselves; and since the sets have different things in them, they aren't the same set obviously.

III. Two special sets

There is a set which contains nothing. It's called the empty set, or the null set, and it's given a special name: **Æ.** (NB: This isn't the number zero; it's the null sign, and in set theory it's reserved for just this one use).

(16) $\{\} = \varnothing$

(17) $| \varnothing | = 0$

There's another set, which contains absolutely everything there is. It's the Universal Set, and the name for it is: **U**. There is some debate about whether there is actually a set containing absolutely everything (including all sets, and so it'd contain itself, which leads to nasty paradoxes). In our discussion of NL semantics, all we'll usually care about is a background set of everything accessible in the discourse, which is usually called "the universe of discourse".

Symbols are starting to proliferate, so here's a summary so far:

Î , Ï is / is not a member of

| | the cardinality symbol; always appears with something in between the two lines, as in | A |, and is read as "the cardinality of set A" (i.e.e the numerical size of set A, i.e. the number of things that are members of A).

lower case letters, late alphabet (e.g., **x, y, z, v, w**..): variables for individual objects, used in the specification of a set by abstraction/description.

{} the set description symbol; always appears with something inside, which names of describes what members the set has.

| , as in { x | x is ...} read as "such that"

capital letters, **early alphabet:** letters which are used as names of specific sets, for convenience. As in e.g.,

C = {w | w is a cat}

capital letters, late alphabet: letters which are used as variables over sets. As in

| X | means "the cardinality of set X (for any set X)".

Notice that the various kinds of descriptions are going to be useful in describing various basic components of natural language. In set theory, we need to consider both specific objects (things we have names for in HL, or use deictic pronouns to refer to), and make generalizations across objects (which we do with the quantifiers of HL). And we need to talk about, in specifying sets, properties that indefinitely large groups of objects have; these will be useful in explicitly describing 1-argument predicates (like adjectives, nouns, and intransitive verbs).

III. Different descriptions of the same set.

What matters to the identity of a particular set is *the actual things that are in it*, not the way we describe them. So, these sets are the same:

(18) {George Washington, Andy Barss}
(19) {the first president of the US, the 2005 instructor for LING 564 at the UA}

Similarly, since ‘3’, ‘three’, ‘III’, ‘the positive square root of nine’, and ‘the only odd integer between one and four’ all are different ways of describing the same thing, these sets are the same:

(20) a. {3}
b. {three}
c. {III}
d. {the positive square root of nine}
e. {the only odd integer between one and four}

IV. Two things that Don’t go into a set description

The **order** in which items are listed in a set description has no relevance to the set. So, {3, 4, 5} is the same set as {4, 3, 5}, and so on for other permutations. (There are formal constructs in which order does matter; we’ll see them when we talk about relations and functions).

Since the description of the object isn’t a member of the set (the object described is), **the same thing cannot be a member of the same set twice** (or more than once). That is, (21) is not a well-formed set description, nor is (22):

(21) {3, 3}
(22) {3, the only odd integer between one and five}

V. Basic Relationships Between Sets

Suppose we have two sets. They might (depending what's in them)

a. have some members in common
b. have no members in common
c. have all members in common
    i. they are the same set
    ii.the aren't the same, but the members of one set are all members of the other

These relations between sets are going to be very important in precisely describing HL. Luckily enough, these relationships are well-understood in set theory, and we have a handy notation for each of them. Below is the repertoire of symbols and illustrations of them. (I'm going to try to avoid using the word "relation", which has a specific technical definition we'll see next time).

A. The Subset Relationship

For two sets X and Y, X is a subset of Y iff every member of X is also a member of Y.

When we just want to assert that X is a subset of Y – whether X is a proper subset, or, perhaps, they are the SAME set – we use the symbol **Í** , as in

$$X \subseteq Y.$$

read as "X is a subset of or equal to Y".

There are two cases to distinguish. One is just the subset relation, as described above. The other is when Y contains at least one member which is not a member of X. In this case, we call X a proper subset of Y, and the symbol used is

$$X \subset Y$$

read as "X is a proper subset of Y".

The subset symbol can also be negated, so that

$$X \not\subset Y$$

means "X is not a proper subset of Y". And

$$X \not\subseteq Y$$

means “X is neither a proper subset nor equal to Y”.

In usual practice, the symbols for subset and proper subset are used most of the time, and the reversed symbols for superset/proper superset (see below) are used less.

B. The Superset Relationship

This is the inverse of the subset relationship. That is, Y is a superset of X just in case X is a subset of Y. Y must have, among its members, everything that is a member of X (and maybe other stuff). The symbol for the subset relationship, written backward, represents the superset relationship:

$Y \supseteq X$ “Y is a superset of X”
$Y \supset X$ “Y is a proper superset of X”

Some examples. Consider these sets:

(23) $A = \{x \mid x \text{ is a cat}\}$
$B = \{y \mid y \text{ is a mammal}\}$
$C = \{z \mid z \text{ is a sofa}\}$
$D = \{w \mid w \text{ is a couch}\}$

Relations between these sets:

(24) $A \subseteq B$
$A \neq B$
$A \subset B$
$B \supseteq A$
$B \supset A$

$C \subseteq D$
$C = D$
$C \not\subset D$

$D \supseteq C$
$D \not\supset C$

VI. Operations on Sets

There are several formal operations which take two sets as the input to the operation), and give back a third set. These include union, intersection, and subtraction.

A. Union

Defines a new set gotten by combining the members of A and B.

Symbol: **È**

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$$

B. Intersection

Symbol: **Ç**

Defines a new set gotten by taking those members that A and B have in common, and forgetting the ones they don't.

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$$

C. Subtraction

Symbol: **-**

Defines a new set, gotten by removing from one set the members it has in common with a second set.

(26) $A - B = \{ x \mid x \in A \text{ and } x \notin B \}$

This is also called the **complement of set B with respect to set A.**

In some versions of set theory, which admit a "set of all objects", there is also the notion of "an absolute complement of a set". The notation for this (for set B) is:

(27) B^{-}

The idea is this: if you start with the set of all things (call it U), and subtract B from it,

(28) $U - B$

what do you get? Well, you get everything that is not a member of B. So,

(29) $U - B = \{ x \mid x \in U \text{ and } x \notin B\}$

The first half of the description (“$x \in U$”) is unnecessary, since x is a variable for things, and U is the set of absolutely all things. So we can simplify this to

(30) $U - B = \{ x \mid x \notin B\}$

And the use of the $\overline{}$ symbol is precisely to indicate the complement of a set with respect to U, so

(31) $\overline{B} = U - B$

That’s pretty much all we need to know about sets to get the theory of language up and running. We will see variants of some of these ideas and relations in predicate logic, which we begin in a week.

Homework Assignment #1. (Please type. Use symbol font for the fancy symbols if you can; if you must write something in by hand, be clear).

Problem 1. Please do exercise on sets in Heim and Kratzer, page 9-10.

Problem 2. What is the cardinality of the sets in (i), (ii)? Write your answer using the cardinality symbol. Also describe in English why you think this is the answer.

(i) $\{\{a,b\}\}$
(ii) $\{\{\varnothing\}\}$

Problem 3: Describe what’s wrong with the set descriptions below.

(a) {x | y is larger than a breadbox}
(b) {George Washington
(c) {George Washington, the first president of the US}