HG4041 Theories of Grammar

Grammar and Processing

Francis Bond

Division of Linguistics and Multilingual Studies

`http://www3.ntu.edu.sg/home/fcbond/`

`bond@ieee.org`

Lecture 7
Location: HSS SR3

Overview

➤ Psycholinguistics and grammar design

   ➤ What grammar has to say
   ➤ What psychological evidence has to say
      ∗ Acquisition
      ∗ Production
      ∗ Comprehension
   ➤ Universals

What does grammar have to do with psychology?

Three ways it could be relevant:

➤ It provides insight into how children acquire language.

➤ It provides insight into how speakers produce utterances.

➤ It provides insight into how listeners understand utterances.

Surface-oriented

Constraint-based

Lexicalist

Chomsky's position:

➤ Grammar represents knowledge of language (**competence**).

➤ This is distinct from use of language (**performance**).

➤ We can draw a strong conclusion about language acquisition, namely, most grammatical knowledge is innate and task-specific.

➤ Serious study of language use (production and comprehension) depends on having a well-developed theory of competence.

Brief remarks on language acquisition

➢ Chomsky's nativism is very controversial

  ➢ It is based on the **poverty of the stimulus** argument, and a model of learning as hypothesis testing.
  ➢ The environment may be more informative than he assumes.
  ➢ There may be more powerful learning methods than he assumes.

➢ There has not been much work on language acquisition using constraint-based lexicalist theories like ours; but

  ➢ Explicit formulation is a prerequisite for testing learning models
  ➢ Our feature structures could model richer context information.

➢ We're neutral with respect to this controversy.

Production and Grammar

➤ Evidence for left-to-right effects

➤ Evidence for top-down planning

Disfluencies are sensitive to structure

Repeat rate of *the* varies with position and complexity of the NP it introduces:

➤ More common in front of complex NPs

➤ More common with prominent NPs
Topic > Subject > Direct Object > Preposition Object

   (1)   *The the book I told the student about on the train*

Production errors are sensitive to syntactic structure

➤ Agreement errors are more common with PP complements than sentential complements: errors like (3) are significantly more common than errors like (2).

    (2)   *_The claim that the wolves had raised the babies were rejected._

    (3)   *_The claim about the newborn babies were rejected._

➤ Why?

    ➤ Speculation: Clauses are their own agreement domains, so people don't mistake an NP in a lower clause as a trigger for agreement

    ➤ Original work: Kay Bock (1980s).

Some high-level sentence planning is necessary, too

(4) *Ich habe dem   Mann, den   ich gesehen habe geholfen.*
     I   have the-dat man   who-acc I   seen   have helped

     "I helped the man I saw"

(5) *Ich habe den   Mann, dem   ich geholfen habe gesehen.*
     I   have the-acc man   who-dat I   helped   have seen.

     "I saw the man I helped "

➤ The choice between *dem* and *den* depends on the choice of verbs several words later.

Interaction of top-down and left-to-right information

➤ Grammar plays a role in production.

➤ Partial grammatical information should be accessible by the production mechanism as needed.

➤ This argues against grammatical theories that involve sequential derivations with fixed ordering.

➤ Our theory of grammar has the requisite flexibility.

Comprehension

➤ Early work tried to use transformational grammar in modeling comprehension

➤ **The Derivational Theory of Complexity**: The psychological complexity of a sentence increases with the number of transformations involved in its derivation.

➤ Initial results seemed promising, but later work falsified the DTC.

Some relevant quotes

➤ "The results show a remarkable correlation of amount of memory and number of transformations"

— Chomsky (1968)

➤ "[I]nvestigations of DTC. . . have generally proved equivocal. This argues against the occurrence of grammatical derivations in the computations involved in sentence recognition"

— Fodor, Bever, & Garrett (1974)

➤ "Experimental investigations of the psychological reality of linguistic structural descriptions have . . . proved quite successful."

— Fodor, Bever, & Garrett (1974)

➤ In particular, they concluded that **deep structures** and **surface structures** were psychologically real, but the transformations relating them weren't.

Evidence for the Psychological Reality of Deep Structures

➤ The proposed Deep Structure for (7) had three occurrences of *detective*, while the proposed DS for (6) had only two:

    (6)   *The governor asked the detective to prevent drinking.*
    (7)   *The governor asked the detective to cease drinking.*

➤ In a recall experiment, *detective* was significantly more effective in prompting people to remember (7) than (6).

Typical Problem Cases for the DTC

(8) *Pat swam faster than Chris swam.*

(9) *Pat swam faster than Chris did.*

(10) *Pat swam faster than Chris.*

➤ The DTC predicts that (8) should be less complex than (9) or (10), because (9) and (10) involve an extra deletion transformation.

➤ In fact, subjects responded more slowly to (8) than to either (9) or (10).

What should a psychologically real theory of grammar be like?

➤ The **deep structure** distinctions that are not evident on the surface should be represented.

➤ The transformational operations relating deep and surface structures should not be part of the theory.

➤ Our information-rich trees include all of the essential information in the traditional deep structures, but without the transformations.

Jerry Fodor claims the human mind is modular

"A module is . . . an informationally encapsulated computational system –
an inference-making mechanism whose access to background information is
constrained by general features of cognitive architecture."

— Fodor, 1985

➤ A central issue in psycholinguistics over the past 20 years has been whether
language is processed in a modular fashion.

Tanenhaus's Eye-Tracking Experiments

➤ Participants wear a device on their heads that makes a videotape showing exactly what they're looking at.

➤ They listen to spoken instructions and carry out various tasks.

➤ They eye-tracking provides evidence of the cognitive activity of participants that can be correlated with the linguistic input.

Non-linguistic visual information affects lexical access

➤ Participants' gaze settled on a referent before the word was completed, unless the initial syllable of the word was consistent with more than one object.

➤ For example, participants' gaze rested on the pencil after hearing *Pick up the pencil* more slowly when both a **p**encil and a **p**enny were present.

Non-linguistic visual information affects syntactic processing

➤ Eye movements showed that people hearing (11) often temporarily misinterpreted on the towel as the destination.

  (11)   *Put the apple on the towel in the box.*

➤ When *on the towel* helped them choose between two apples, such misparses were significantly less frequent than when there was only one apple.
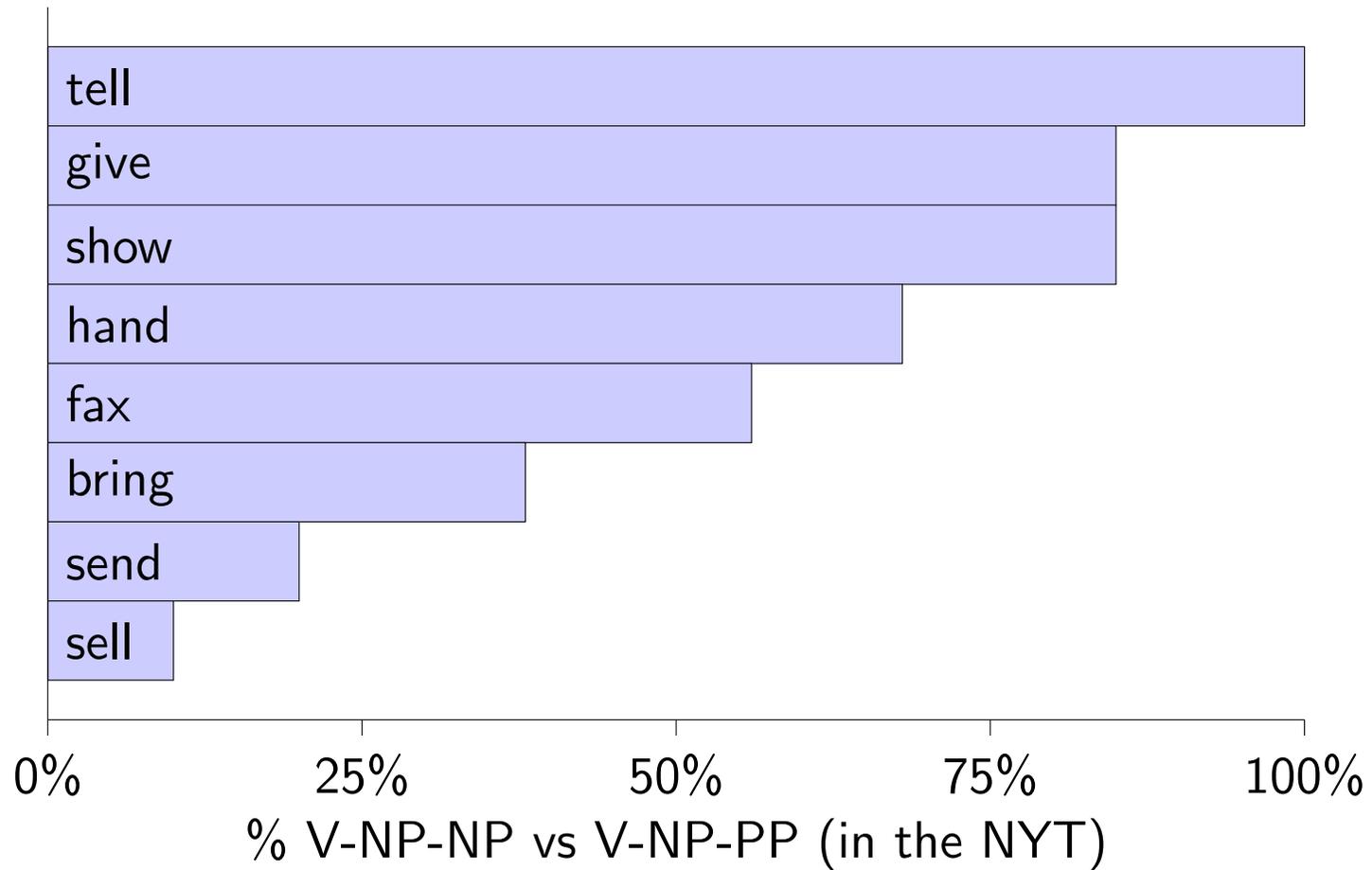
General Conclusion of Eye-Tracking Studies

➤ People use whatever information is available as soon as it is useful in interpreting utterances.

➤ This argues against Fodorian modularity.

➤ It argues for a model of language in which information is represented in a uniform, order- independent fashion.

Speakers know a great deal about individual words

➤ Individual lexical items have many idiosyncrasies in where they can occur, and in where they tend to occur.

➤ For example, the verb *behoove* occurs only with the subject *it* (and only in certain verb forms), and the verb *beware* has only the base form.

➤ We also know that the transitive use of *walk* is much rarer than the intransitive.

Different verbs favor different COMPS lists

Horizontal bar chart:
- tell: ~100%
- give: ~85%
- show: ~85%
- hand: ~68%
- fax: ~57%
- bring: ~38%
- send: ~20%
- sell: ~10%

X-axis: 0%, 25%, 50%, 75%, 100%

% V-NP-NP vs V-NP-PP (in the NYT)

Lexical biases influence processing

➤ Wasow et al. ran a production experiment to test whether ambiguity avoidance would influence speakers' choice between (12) and (13):

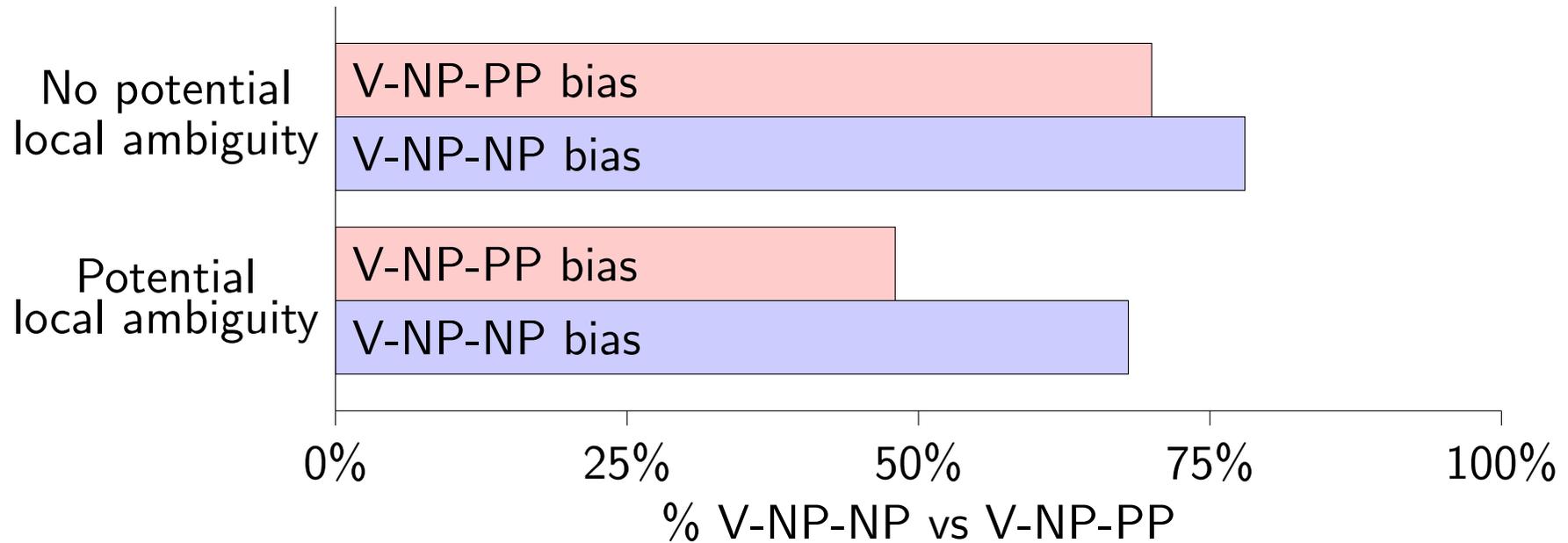(12) *They gave Grant's letters to Lincoln to a museum.* NP-PP
(13) *They gave a museum Grant's letters to Lincoln.* NP-NP

➤ Ambiguity avoidance predicts that you should prefer (13)

➤ Lexical bias of the verbs turned out to be a significant predictor of which form speakers used (and ambiguity avoidance turned out not to be).

Experimental Method

➤ Speaker and Listener sit next to each other. Speaker can see a screen.

➤ Speaker silently reads a sentence shown on the screen
   A museum in Philadelphia received Grant's letters to Lincoln from the foundation.

➤ The sentence disappears from the screen.

➤ Listener asks a question:
   What did the foundation do?

➤ The speaker answers the listener's question.
   The foundation gave . . . . the museum, um, Grant's letter's to Lincoln.

➤ The listener records which kind of response on a list (from two choices).

Experimental Results on Local Ambiguity

No potential
local ambiguity

V-NP-PP bias

V-NP-NP bias

Potential
local ambiguity

V-NP-PP bias

V-NP-NP bias

0%   25%   50%   75%   100%

% V-NP-NP vs V-NP-PP

➤ Arnold, Wasow, Asudeh & Alrenga (2004: *Journal of Memory & Language*) re-ran the experiment with slightly better methodology and found an even stronger reverse ambiguity effect.

A psychologically real grammar should be lexicalist

➤ Early generative grammars downplayed the lexicon.

➤ Now, however, the importance of the lexicon is widely recognized.

➤ This aspect of grammar has been developed in greater detail in HPSG than in any other theory.

➤ It would be easy to add frequency information to the lexicon, though there is debate over the wisdom of doing so.

➤ Frequency is currently recorded as part of the **parse ranking model**s which select the most plausible out of all possible interpretations.

Conclusion

➤ Grammatical theory should inform and be informed by psycholinguistic experimentation.

➤ This has happened less than it should have.

➤ Existing psycholinguistic evidence favors a constraint-based, lexicalist approach (like HPSG).

Universals?

➢ **Principles and Parameters (P&P)**: attempts to relate multiple typological properties to single parameters (top-down).

➢ **Grammar Matrix**: attempts to describe many languages in a consistent framework and then takes stock of common constraints (bottom-up).

Universals?

➢ Case constraint

➢ SHAC

➢ Binding theory

➢ Head-complement/-specifier/-modifier

➢ Head Feature Principle

➢ Valence Principle

➢ Semantic Compositionality Principle

➢ ...

These are candidate universal facts about language 29

Overview

➤ Psycholinguistics and grammar design

    ➤ What grammar has to say

    ➤ What psychological evidence has to say
        ∗ Acquisition
        ∗ Production
        ∗ Comprehension
    ➤ Universals

P1: Constant Rules

The Singular Noun Lexical Rule, the Non-3rd-Singular Verb Lexical Rule, and the Base Form Lexical Rule are all inflectional lexical rules ($i\text{-}rule$) which have no effect on the shape (i.e. the phonology) of the word.

A. Explain why we need these rules anyway.

B. Each of these rules have lexical exceptions, in the sense that there are lexemes that idiosyncratically don't undergo them. Thus, there are some nouns without singular forms, verbs without non-third-person singular present tense forms, and verbs without base forms. List any you can think of.

Questions

Q: On page 294, constraint-based lexicalism is mentioned, and it is noted that different speakers have differing language use (and thus grammar) due to the different cultural, socio-economic and political influences they go through. My question is, if that is they case, how do we, when doing lexical data etc, prevent ourselves from imposing our own grammatical judgements onto others, which may lead to data and the final form being (wrongfully) analysed?

A: We either model our own grammar (for which we have our own judgments) or base our grammar on a corpus (and try to match the attested data). In practice we normally do both: use our judgment and check against corpora.

Q: So, for a theory of grammar to be deemed 'realistic', it needs to be able to tick as many check boxes in explaining as many features of linguistic performance?

A: Yes.

Q: The theory that the chapter covers abides by ensures that the utterances are simplistic in nature. How would the theory hold up to more complex usage of language?

A: The main point is that the mechanisms are compatible with what we know about performance. More complex grammars exist and they cover a surprising amount of English. So far, this also appears possible for other languages.

C: Thanks for the reminder. I actually don't have questions for this week's reading (as of yet). The only comment I have so far is that I really like how the authors have used major features of HPSG to evaluate critique innate grammar.

Q: Oh maybe I do have a question but it's not directly related to this week's lecture. If Chomsky's theory is so controversial and increasingly doesn't seem to reflect how we use/learn language, why are we learning so much about him?

A: Because it is still the most widely taught in the US (and possibly in the world).

Q: In 9.5.1 (surface orientation), the book states that that word order matters in sentence processing, as sentences are built in the order of the information that is given, but in 9.5.2 (Constraint-Based Grammar) it seems to contradict it by saying that word order only informs the order of 'mental operations' listeners use to understand sentences.

Aren't these 'mental operations' conducted simultaneously as more information is revealed in a sentence, rather than in any given order? Why is there a distinction between accessing syntactic and semantic information, when both seem relevant to the features of a given word?

A: I didn't see these as contradictory — the mental operations are part of the parsing. The emphasis in 9.5.2 is on the integration of non-linguistic world knowledge, while 9.5.1 is on the timing of the integration.

Q: Using what we have learnt, how would we analyze the following sentences?
(i) "Should be can lah."
(ii) "It should be alright."

A: I am not sure if this is a question or a suggestion for a project. Both involve auxiliaries (which we have yet to do, but we treat as verbs that take a VP complement) and the first also has Singapore-specific verb forms and sentence final particles.

Q: What is the significance of the Lexical Rules? What purpose do they serve, with respect to and in comparison with the rules we've learnt in the past eight chapters? I'm a little confused about the larger picture of this chapter when I read these rules.

A: They allow us to capture generalizations in the lexicon

Q: What is the function of the slash sign before the list features? e.g. SPR / <>
I think you mentioned it in lecture but I couldn't really understand it there and then.

A: It marks defeasible constraints

Q: Could you explain lexemes further? Why do we have this term, and how is it of significance (instead of 'words')?

A: Lexemes are not fully specified, whereas words are complete (they match what actually appear in utterances).

Q: Assuming we do not run through the list of conditions satisfied by a good tree structure during lecture on Tuesday, could we run through this list to get a proper understanding of the conditions?

A: OK

Q: Can you explain what it means to have "no crossing branches" (as on Page 291)?

A: The lines joining nodes cannot cross each other. I will draw this :-).

Q: I'm not sure if this is asking too much, but from my sensing of the ground (aka my classmates), I think we need some practice from the basics. Most of the problems seem to pose slightly more sophisticated questions, and it seems that we've been going at quite a blistering pace in terms of covering the contents.

A: It may help to read a few chapters behind and ahead each week.

Q: One challenge I find in this module is that the grammar (of both words and phrases) (and now lexemes) is analysed differently in different chapters (probably because as it progresses, new things are introduced). So it would help if we get some time to piece the puzzle of HG4041 before we go into the dark alleys of the final assignment.

A: Fewer changes are introduced from now on (new things are added, but there is very little change), so I hope there will be time to be more comfortable.

Q: With regards to this week's readings, I would like to ask an actual question but I do not actually know what all the various symbols represent. (For example, Aatom = Apol U Agr.atom U Amode U Areln ?????)

A: U is Union, and the other things are described just below in the text (but see below)

Q: Also, I have reread pg 290 over and over again but still cannot make any sense of it.

A: It is ok to skip the squiggly bits.

Q: Week 7 Question: If this Type Hierarchy Model were to be applied on another language with grammatical gender, do we add additional gender features under the determiner, noun, or both?

For example in Spanish, 'The female/male lawyer' is written as follows:
(1) El Abogado 'The male lawyer'
(2) La Abogada 'The female lawyer'

From the examples, we can see that both the determiner and the noun is conjugated accordingly to the gender (i.e. article el and suffix -o for masculine nouns and article la and suffix -a for feminine nouns).

A: What do you think? What happens with larger structures?

Q: In 9.4.1's point about the relevant non-linguistic information, could it be understood sort of using the the universal subject selection rule? i.e. A more agentive noun is likely to be the subject for the main VP, thus in the case of the horse, it makes interpretation of the sentence more ambiguous?

A: Yes, agentivity is almost certainly one of the factors that helps us choose the correct interpretation.

Q: FORM feature

It seems, contrary to what we were lead to believe through the book, that FROM is now a feature of all types of pos (including, for example nouns and adjectives). [at least this is what the hierarchy presented in page 273 suggests]

This might come in handy to re-establish the coordination rule presented in page 280, since the necessity of having a feature FORM was barring NP coordination since the last chapter.

Nevertheless, in the nouns given as examples, do not carry information on FROM. What value would they be? noun for nouns, adjective for adjectives, and so on?

[Why does the book keep sneaking unexplained things like this?]

A: nform, afrom, . . . (p250). But in fact, on page 250 it explicitly discusses FORM for different POS and implies that all POSs have it (or at least anything

that can be coordinated). It is explained, and if you look up FORM in the index it leads you write to it.

Q: Constant Lexeme Lexical Rule
As presented in page 282, the output of this lexical rule does not look like a proper lexeme. Beyond the fact the LIST[FIRST, REST] have not been explained, why is this rule's OUTPUT [FIRST 1], and not a normal lexical entry like the others?

A: This rule only adds one constraint, that the INPUT and the first thing in the list of the INPUT and OUPUT is the same: that is that there is no change in the word form, everything else is inherited from above.

C: Following the discussion whether the formalisms introduced here specific to language or application to other general intelligences, it seems (to me) that Unification of AVMs could be applied to binary logic, memory and, for example,

to model congruency of beliefs (based on type hierarchy and and unification). E.g.

human is-a thing non-human is-a thing

If robot has type non-human (and we assume no multiple inheritance, in this case), than a system of beliefs would fail any reasoning that could postulate robots with features of humans (for features assigned only to humans, e.g. sentient-being).

Q: Do you know of any fun work been done on these topics for unification?

A: Yes, there is lots in the literature of default inheritance. Also object-oriented programming.

Q: Shouldn't oughta wordnet be this knowledge repository? [here would be where defaults would work nicely! E.g. assume the real world, unless told otherwise (i.e.

cows don't have wings unless stated otherwise)]

A: This is what ontologies try to do, and it turns out that is hard to get all the details right (but well worth attempting).