Iniernational Journal of English Studies

IJES

UNIVERSITY OF MURCIA

www.um.es/engphil/ijes

L2 Evidence for the Structure of the L1 Lexicon

CHARLES REISS*
Concordia University

ABSTRACT

In this paper I suggest that L2 research could provide answers to questions concerning the structure of L1 grammars that cannot, as a matter of logic, be answered by only examining L1 data and intuitions. In other words, L2 data from an individual P can provide 'external' evidence bearing on the structure of P's L1 grammar, $L1_p$. This type of evidence will be particularly welcome where competing theories of $L1_p$ are extensionally equivalent where they generate the same output representations. I am proposing, therefore, that L2 research need not restrict itself to maintaining consistency with work in theoretical linguistic modeling. Instead L2 research can itself make unique contributions to the general theory of grammar.

In addition to potentially leading to fruitful results, the issues that provide the background to the discussion warrant examination on other grounds, since they help clarify the goals of linguistic research and the compatibility of various frameworks of linguistic theory with their own stated goals.

KEYWORDS: homophony, ambiguity, underspecification, external evidence, Richness of the Base.

* ***Address for correspondence:*** Charles Reiss, Linguistics Program, Concordia University, 1455 de Maisonneuve W., Montreal, Quebec H3G 1M8, Canada. E-mail: reiss@alcor.concordia.ca

I. THREE PROBLEMS

From the linguistic viewpoint the problem facing a child is to acquire the grammar determined by the innately given language faculty' (Universal Grammar) in conjunction with the input provided by the environment, also referred to as 'experience'. It is perhaps misleading to call this a 'problem' since all evidence indicates that the child learns language as effortlessly as it grows hair and gets taller (as pointed out over the years by Chomsky). However, I will refer to the acquisition of a grammar as the HUMAN'S PROBLEM as a rhetorical convenience to compare it with two other processes.

These two other processes correspond to two distinct research programs carried out by professional linguists. However, we shall see that linguists, themselves, are not always clear about which of these tasks they are engaged in. One topic of research can be called the ARTIFICIAL INTELLIGENCE PROBLEM, and it belongs to sub-branches of linguistics like Natural Language Processing and Speech Recognition. Such work attempts the simulation of human intelligence in recognizing and processing linguistic data without regard to whether or not, the model matches the computational methods used by humans. A model that perfectly matched the input-output mappings of a human in some domain of linguistic behavior would be called 'weakly equivalent' to the system actually used by the human. Such a model will generate the same utterances as the human, say, but may do so using completely different algorithms. In such a case we can also say that the human's grammar and the simulation grammar are EXTENSIONALLY equivalent in that; they generate the same surface pattems. They are not INTENSIONALLY equivalent, in that they arrive at these pattems via different algorithms, and perhaps encode different types and levels of representation.

The third issue to discriminate corresponds to the research program of generative linguistics in the Chomskyan mentalist tradition. For convenience, but perhaps at the risk of alienating linguists working in other traditions, I refer to this research program as the LINGUIST'S PROBLEM. The Linguist's Problem involves figuring out which mental grammar an individual has intemalized using the typically insufficient and indeterminate dataavailable. In other words, the Linguist's Problem is to figure out what the solution to the Human's Problem is. The purpose of this paper is to suggest a way of enriching the generative linguist's data set. If valid, this source of data will contribute in a modest way to progress on the Linguist's Problem. To restate things in a slightly different fashion, the (generative, mentalist) linguist's constructed grammar is intended to be the best hypothesis currently available conceming the nature of the solution to the Human's Problem.

We summarize in (1) the discussion thus far:

(1) Three Problems that need to be distinguished

• The *Human's* Problem: Acquiring the grammar determined by UG + Experience.

• The *AI* Problem: Simulation ofhuman intelligence without regard to whether the model proposed matches the computational methods used by humans (weak equivalence). (See Pylyshyn 1984, Chapter 4.)

• The Linguist's Problem: Using insufficient/indeterminate data, figuring out which grammar a human acquires.

It is clear that the tradition of using the term 'grammar' in a systematically ambiguous fashion to refer both to the knowledge state of an individual's language faculty and to a linguist's model of that knowledge is potentially confusing. In the former usage a 'grammar' is taken to be an object in the world (the solution to the Human's Problem), whereas the latter usage implies a possibly incomplete and inaccurate model of such an object (the current solution to the Linguist's Problem).

Quine (1972) claimed that it is incoherent to attempt to choose among competing grammatical models that are extensionally equivalent, since, by definition they are indistinct with respect to the set of utterances they generate. However, Chomsky (1986) argues that Quine's pessimism is unwarranted. First, the object of study within the Chomskyan program is 'I-language', the system of knowledge internalised in individual minds/brains. Thus, even if we cannot determine the exact form of an individual's I-language, this does not imply that there is not a coherent answer to the question of which grammar it is. Our ability to find the correct answer, and the existence of a correct answer, are logically independent issues.

Furthermore, the choice among extensionally equivalent grammars may be broachable not only in principle, but in practice, as well. As Chomsky points out, our theories of differentL1 grammars should be mutually constraining, since, although partially determined by experience, each is a development from the same initial state (S,) of the language faculty (UG):

> Because **evidence from Japanese can evidently bear on the** correctness **of a theory of SO, it can** have **indirect-but very powerful-bearing on the choice of the grammar that attempts to characterize the** I-language **attained by a speaker of** English.
>
> ***Chomsky (1986: 38)***

So, not only is the question of the linguist's choice among competing models of grammar a valid one in principle, but there are empirical facts that can bear on this choice.

To make the discussion more explicit, consider the following idealized situation. Imagine we could model the structure and acquisition of Japanese in two competing ways, involving two theories of S„ UG, and UG,, and two resulting models of the adult grammars based on these

models of S_n, J_α, and J_β, respectively. Now imagine we could similarly model the structure and acquisition of English in two cornpeting ways, involving two theories of S_n, UG, and UG_n, and two resulting models of the adult gramrnars based on these models of S_n, E, and E_δ, respectively. At this point, the best theory of the hurnan language faculty would be one that posited UG, as the correct model of S_n, and the associated J_β and E, as the correct models of the I-languages of Japanese and English speakers. The study of the structure and acquisition of English and the study of the structure and acquisition of Japanese should be mutually constraining. The goal of linguistics is not rnerely to simulate Japanese and English type output. However, we will now see that some current theoretical work is tending in this direction, apparently unintentionally.

II. RICHNESS OF THE BASE AS A CONFUSION OF GOALS

It appears that Optirnality Theory (Prince & Smolensky 1993, McCarthy & Prince 1993 Kager 1999) has tacitly, and perhaps inadvertently, rejected the distinction between the three Problerns in (1) by adopting the principie of RICHNESS OF THE BASE (RoB) as a central tenet of the theory.

According to Kager (p. 19), RoB dernands that "no specific property can be stated at the level of underlying representations". There are no restrictions against certain sequences of segments or against certain feature cornbinations, in fact there are no morpheme structure constraints (MSC's) of any kind.

Kager (p. 31-32) shows that, using any combination of nasal and non-nasal vowels in underlying forms, a single OT constraint ranking for English could generate the correct output forms [sæd] 'sad' and [sæ̃nd] 'sand' . That is, the ranking produces the right output for any of the lexicons in (2):

(2) Ranking: $*V_{ORAL}N >> *V_{NASAL} >>$ IDENT-IO(nasal)

	INPUT		**OUTPUT**
LEX1:	/sæ̃d/ & /sæ̃nd/	>	[szd] & [sæ̃nd]
LEX2:	/sæd/ & /sæ̃nd/	>	[sæd] & [sæ̃nd]
LEX3:	/sæd/ & /sænd/	>	[sæd] & [sæ̃nd]
LEX4:	/sæd/ & /sænd/	>	[sæd] & [sæ̃nd]

The most highly ranked constraint is violated when an oral vowel occurs directly before a nasal consonant. The next constraint is violated by the appearance of nasal vowels in output forms. The lowest constraint demands input-output identity for the feature [nasal]. This is typical of how OT generates allophonic variation: a context-sensitive markedness constraint is ranked above a potentially conflicting context-free rnarkedness constraint; and both are ranked above a relevant

faithfulness constraint. Without being tied to a unique view of what the lexical items are, the proposed ranking generates the right results.

It is not completely clear what Kager's data is meant to show. One possibility is that the grammar licenses all of these derivations, they are all equally 'psychologically real'. Another possibility is that the data is meant to show that several extensionally equivalent choices exist, and that there are no grounds for choosing among them. In other words, it is unclear whether a theoretical point conceming the nature of grammars is being made, or whether the point is a metatheoretical one concerning the limits of our knowledge. In fact, Kager's discussion is a useful demonstration of an interesting mathematical property of OT grammars, but I will argue that it is psychologically uninteresting.

Before proceeding to the relevance of RoB to the three Problems, it is worth pointing out that it is odd to consider RoB as apart of OT. RoB actually just lists properties that other theories may have, but OT does not. There is an infinitely long list of properties that any particular version of OT does not have. For example, none have lust or interest rates, but we clearly do not want to encode this infinite list explicitly in finite human minds. This problem of defining grammars in negative terms is discussed further in Reiss (2000).

However, there are more insidious aspects of RoB. These include explicit appeals to this non-principie, as well as the fact that some OT analyses are dependent upon RoB not being valid. That is, they only make sense if MSC's and constraints on underlying phoneme inventories are included in the grammar!

McCarthy (1999a:6) invokes RoB to avoid selecting a single underlying representation for a given morpheme: "with faithfulness bottom-ranked, the choice of input [among three alternatives] doesn't matter, since all map to [the same surface form]. So there is no need to restrict the inputs." McCarthy has solved the Artificial Intelligence Problem, since his constraint ranking generates the same (correct) output representation for all three of the input forms he considers. However, generative linguistics is not Artificial Intelligence, and McCarthy turns his back on the Linguist's Problem by claiming that the choice between competing input forms "doesn't matter". The Human leamer must have stored something, and it is the phonologist's job to tell us what.

We could, following McCarthy, also posit that there may be *bananas*, not representations of bananas, but actual bananas, in the underlying representation of words. Since the language faculty presumably cannot assign a pronunciation to a banana, we never get direct evidence that the banana is there. It seems perverse to leave the door open to the possibility that it 'doesn't matter' whether we choose the bananaful or the banana-free representation, but this is basically what McCarthy is doing.

Generative grammar as cognitive science is interested in the solution to the Human's Problem: if McCarthy is right in saying that the choice among underlying forms does not matter, then cognitive science does not matter. I assume that we can reject this possibility. What

McCarthy is doing is confusing various issues in advocating no restrictions on inputs.

There is no question of 'restricting' the inputs in the sense of positing MSC's as part of the grammar, but rather a question of figuring out which inputs the learner constructs given the observed data. It is something of a perversion of terms to label our hypothesis about what the LAD does a 'restriction', when in fact we mean 'arriving at a solution, given data and a learning algorithm'.

A supposed benefit of incorporating RoB in a theory of grammar is that the surface inventory (both in phonology and morphosyntax) is then taken to be predictable from the constraint ranking. This is taken as support for the notion that differences among languages reduce to differences in ranking. Consider two more discussions of RoB in the OT literature:

> The set of possible inputs to the grammars of all languages is the same. The grammatical inventories of languages are defined as the forms appearing in the structural descriptions that emerge from the grammar when it is fed the universal set of all possible inputs. Thus, systematic differences in inventories arise from different constraint rankings, not different inputs. The lexicon of a language is a sample from the inventory of possible inputs; all properties of the lexicon arise indirectly from the grammar, which delimits the inventory from which the lexicon is drawn. There are no morpheme structure constraints on phonological inputs, no lexical parameter that determines whether a language has ***pro.***
>
> ***Tesar & Smolensky (1998: 252)***

> We can also see that it is inevitable that light do exists in [English], given the constraint rankings.
> *Grimshaw* ***(1997: 387)***

It is apparent that RoB gets things exactly wrong: if we accept an OT model of grammar, then the inventory present in the PLD will determine how the learner ranks the constraints. All learners start out with the same constraints in the same ranking relationships (under an OT view of UG). The way learners come to acquire different grammars is due to the differences in their experience-the inventories they are exposed to. This confusion of cause and effect reflects more general problems with existing models of acquisition in OT discussed by Hale & Reiss (1998).

If we can now agree that we care about making hypotheses about what constitutes knowledge of language, and how that knowledge comes to be instantiated in human minds —and not just what can make a system act like it has knowledge of language— then we are ready to tackle the vexing problem of HOMOPHONY in natural language.

III. HOMOPHONY IN PHONOLOGY

The practice of phonological and morphological analysis involves modeling two kinds of situation: (a) deriving surface distinctions from identical substrings of underlying representations;

and (b) demonstrating that identical surface strings can correspond to underlyingly distinct representations. In phonology, these two aspects of analysis are fairly well understood. The first (a) is achieved by positing context-sensitive processes which selectively affect parts of representations in accordance with the contexts in which they appear, and the second (b) by positing neutralization processes.

(3) Phonological analysis

a. *One-to-many mappings*—deriving surface distinctions from identical inputs-context sensitive processes.

b. *Many- to-one mappings*—4emonstrating that identical surface strings can correspond to underlyingly distinct representations-neutralization processes (phonologically-based homophony)

After illustrating these two phonological patterns, we will turn to their correspondents in morphology.

III.1 Phonological relationships in Old Icelandic

A simple case of the derivation of a surface distinction from a single underlyingly representation is provided by the Old Icelandic noun paradigms in (4).

(4) Assimilation of *-r* to coronal sonorant (Reiss 1994)

case	'home'	'stone'	'wagon'
NOM	/heim-r/ → *heimr*	lstein-rl → *steinn*	/vagn-r/ → *vagn*
ACC	/heim-Ø/ → *heim*	/stein-Ø/ → *stein*	/vagn-Ø/ → *vagn*

Masculine nouns of this class bear the suffix **-r** in the nominative singular and no overt suffix in the accusative singular: *heim-r/heim* 'home nom./acc. sg.', *arm-r/arm* 'arm nom./acc. sg'. The masculine singular marker assimilates to a preceding coronal sonorant under certain well-defined conditions (see Reiss 1994 for an analysis), so that the nominative singular of the stem *stein-* 'stone' is steinn and the accusative is *stein.* We thus see that the surface distinction between nominatives with *-r* and nominatives with gemination of the final coronal sonorant can be derived phonologically from a single underlying suffix.

Compare this situation to that of stems ending in a consonant cluster of increasing

sonority, like vagn- 'wagon' andfugl- 'bird'. Here the nominative suffix is deleted. Like the case of gemination just discussed, this deletion process is purely phonological. The nominatives are vagn andfugl, not **vagnr*, **fuglr* or **vagnn, *fugll*. In other words, the nominative ends up being identical with the accusative in its overt phonetic realization: the nominative and accusative are homophonous. However, common practice, and common sense too, tell us that there are two different sets of morphosyntactic features corresponding to each string —one marked nominative (e.g., *vagn*[NOM]) and one marked accusative (*vagn*-[ACC]). The surface homophony, or ambiguity, in such a situation is derived.

Instead of accepting the phonological account of the homophony of the nominative and accusative of vagn-, we might propose that masculine nouns take the nominative ending -r, except in the case of stems ending in certain clusters. However, it then would be treated as an accident that these clusters can be described as a natural class using universally required linguistic primitives. And we might predict that any random list of stem shapes could be similarly exceptional. The ability to capture generalizations is a standard argument against adopting such a hypothesis.

IV. HOMOPHONY IN MORPHOLOGY

In morphology, situation (a), in which surface distinctions between strings which overlap in meaning or morphosyntactic distribution, but which cannot be derived from the 'lower' phonological level, is attributed to one of two causes. Either we are looking at a case of root suppletion (ai), or else the overlap in meaning results from the concatenation of one constant element with different lexical items in different contexts (aii).

(5) Morphological analysis

a. One-to-many mappings —deriving surface distinctions from identical inputs

i. root suppletion: *go/went*

ii. context dependent suppletion of inflectional morphemes: Hungarian hajó-k 'boats' hajó-i-m 'my boats'

b. Many-to-one-mappings demonstrating that identical surface strings can correspond to underlyingly distinct featural representations, or alternatively, demonstrating that the inputs are actually identical-the problem of **morphological homophony.**

As noted, example (5ai) represents typical root suppletion. The present and past tense forms of this verb are not synchronically relatable by productive morphological or phonological processes. Example (5aii) is in essence identical to the preceding one, and is separated here only because such cases are not typically referred to as suppletion. The plural marker on Hungarian nouns is -ok (or harmonic variants) on nouns that are not marked with a possessive suffix, but the plural marker is -i- when the noun also bears a possessive suffix. Thus, the choice of marker depends on the morphological context. The difference is that unlike *went,* Hungarian *-i-* does not express in an unanalyzable unit a combination of root and inflectional features.

The principles of linguistic analysis in situation (5b) are most difficult to explicitly characterize. This is the situation where potentially distinct morphosyntactic structures are realized by identical phonological strings. Often the crucial question is actually whether or not the phonological constant does correspond to multiple morphosyntactic structures. In other words, our understanding of the treatment of potential HOMOPHONY is unacceptably vague. The present paper does not attempt a thorough treatment of morphological homophony. Instead, 1 will define this type of homophony, demonstrate that it in fact exists, and then suggest how L2 data could help us to recognize it.

An example, again from Old Icelandic, will help. This language distinguishes up to four cases overtly in noun paradigms. The nominative and accusative have already been mentioned. In addition, there are overtly marked genitives. A full paradigm for the masculine noun 'home' is given in (6).

(6) A masculine noun paradigm

	SING.	PLUR.
NOM.	*heimr*	*heimar*
GEN.	*heims*	*heima*
DAT.	*heimi*	*heimum*
ACC.	*heim*	*heima*

In contrast to the masculines, Old Icelandic never distinguishes nominative and accusative in neuter nouns, either in the singular or the plural, as illustrated by the paradigm for the noun meaning 'ship' in (7). In this particular case, singular and plural forms are identical, but this is not the case for all neuter nouns.

(7) A neuter noun paradigm

	SING.	PLUR.
NOM.	*skip*	*skip*
GEN.	*skips*	*skipa*
DAT.	*skipi*	*skip*
ACC.	*skip*	*skip*

The question now arises whether Old Icelandic has, for example, just one singular form skip that is neither nominative nor accusative but shares the features common to both usages, or two separate forms, one nominative and one accusative, that happen to be homophonous. The first possibility will be referred to as the theory of ***vague*** or ***general*** representations. The vague representation would be something like this: *skip*–[SG NOUN ANIMATE]. Implicit in this view is a theory of underspecification —surface forms may be partially underspecified for the features of the morphosyntactic context into which they are inserted. The second possibility will be referred to as the theory of ***ambiguous*** or ***homophonous*** representation. The string skip would be part of two distinct representations: *skip*–[NOM SG NOUN ANIMATE] and *skip*–[ACC SG NOUN ANIMATE]. The first three terms ***(vague, general, ambiguous)*** are discussed by Bresnan (1999, q.v. for references) who adopts the stand that grammars make great use of vague representations, This choice allows Bresnan to incorporate her underspecified representations into a theory of formal markedness.

According to Bresnan (1997), the use of ***vague*** or ***general*** interpretation means that unspecified features are necessarily absent from a representation; and ***ambiguous*** interpretation refers to a situation with a set of (potentially) overlapping structures which differ with respect to certain features. I adopt the term *ambiguous* in deference to tradition, but in fact the term involves a mixing of levels of analysis. A phonological string is called 'ambiguous' if it corresponds to more than one morphosyntactic feature structure. Taken as a linking of a phonological representation and a morphological feature structure, lexical items cannot be ambiguous. Instead we are just looking at homophony. So 'ambiguous' can be read as a synonym for ***homophonous*** in the following discussion[2].

A third logical possibility exists, in addition to vague and ambiguous representations. This view is less often considered in the morphological literature and it holds that there may exist a single form which is specified for features that are compatible with all contexts in which a string appears. For example, the string ***skip*** corresponds to a feature structure which is ***both*** nominative and accusative. This third alternative will not be considered below, but see Dalrymple and Kaplan (2001) for interesting arguments in favor of this view. For now, we concentrate on the first two possibilities in the Icelandic case and in similar cases from a variety of languages.

Note that the three views sketched here are not mutually incompatible. Languages could possibly contain representations of all three types. However, the task of the linguist will be to

discover the principles governing the nature of a representation in each particular instance. Ultimately these principles should be defined in terms of the learning path of language acquisition. In this paper, I am concerned with merely identifying the problem and suggesting a potentially useful source of relevant data.

Before proceeding, it is worth distinguishing an issue which is related to, but not identical to the issue under discussion. This is the issue of phonologically null morphemes. For example, we could imagine the existence of a stored lexical item ***skip*** which has no case or number features, and only receives them in combination withpotentially phonologically null inflectional affixes. Alternatively, we could imagine that a form ***skip*** is stored with inherent case and number features matching those contexts in which the form is inserted. For our purposes, we can often sidestep this issue, since our primary concern is to compare the morphosyntactic features associated with actually occurring pronounced words. This discussion must be clearly fine-tuned for compatibility with different theories of morphology.

V. TWO LOGICAL EXTREMES

V.1. Radical vagueness

An extreme version of the theory of vague representations, which is probably not explicitly held by any scholar, could be formulated as follows:

(8) Radical vagueness

> There is no homophony (other than that which can be derived phonologically). In a given language, a single underlying phonological representation (input to the phonology, UR) Σ corresponds to a single morphosyntactic feature description which subsumes the description of all the morphosyntactic environments in which Σ can appear.

According to radical vagueness, there is just one Old Icelandic word (that is, one featural representation) ***skip*** that denotes nominative and accusative for both singular and plural. This seems plausible enough, but radical vagueness has other implications.

It is obvious that if we admit the existence of phonologically null morphemes, then radical vagueness will require that each language contain only a single one. This is because phonologically null morphemes will always have the same (null) underlying representation, and any two would thus be homophonous, ***contra*** radical vagueness.

Even if we exclude the case of phonologically null morphemes, we can still demonstrate that (8) is untenable. It would require that the phonological string found in English ***well*** correspond to a single lexical entry. In other words, the noun in a ***deep well*** and the adverb in *I*

sing well would have to be stored with a single vague set of rnorphosyntactic features that were compatible with both uses. The existence of clear, accidental homophony in natural language, as well as the fact that it is unlikely that appropriate featural representations could be constructed for a single lexical itern like *well* in this rnodel, rnake the theory of radical vagueness untenable.

V.2. Radical ambiguity

The opposite logical extreme frorn radical vagueness would be radical ambiguity. Instead of the lower lirnit of zero arnbiguity and full vagueness, this theory adopts the upper lirnit of ambiguity (homophony) allowed by UG. Then, sornewhat pretheoretically, we can state the doctrine of radical arnbiguity as in (9).

(9) Radical ambiguity

> If there are *n* morphosyntactic contexts in which a string Σ appears which can be distinguished using the set of all morphosyntactic features provided by Universal Grammar, then Σ is n-ways ambiguous; that is, Σ corresponds to *n* (listed or derived) lexical iterns.

Under radical ambiguity, the so-called nominative/accusative of Old Icelandic neuter nouns actually corresponds to distinct forms, a nominative one and an accusative one. Informally, there are two singular words *skip*, not one, and two plural forms, for a total of four homophonous forms.

However, radical ambiguity does not stop here. The theory demands that *any humanly possible* distinction is potentially encoded. We can use English words to illustrate. The distinction between DUAL and PLURAL is never encoded overtly on English nouns. However, UG provides us with these features. Therefore, the proponent of radical ambiguity would require that the string *cats* correspond to *at least* two feature structures, one marked as DUAL and one as PLURAL. In fact, there would be rnany, rnany more possible feature structures, including ones to encode different definiteness and case distinctions.

V.3. Discussion

It rnight be tempting to make Radical Vagueness slightly more plausible by factoring out clearly accidental homophony of lexical iterns like *well, knight/night* or *(to) fly/(a) fly)*, assuming we could figure out how to do this in principled fashion. However, in the next section I show that there must be even more homophony/ambiguity than just these clear cases. This discussion rnight then tempt us to embrace Radical Ambiguity. A moment's reflection should show that no

language internal evidence can lead us to reject Radical Ambiguity —we can always construct extensionally equivalent grammars that make use of various amounts of ambiguity. English will ***sound*** the same whether we model it with one or more lexical items ***you.*** At the end of the paper we will see how L2 data can bear on these issues, and we will find arguments that the truth lies somewhere between the two logical extremes.

VI. ARGUMENTS FOR THE NECESSITY OF HOMOPHONY

A reasonable goal of morphological analysis is to reduce the number of lexical items ('morphemes' in a traditional sense) needed to account for the surface forms of a language. This goal entails demonstrating that a single form fulfills functions in the grammar which can be distinguished by the analyst. In practice, however, many morphologists recognize the existence of homophonous (ambiguous) forms. For example, the past tense of ***hit*** must contain a phonologically null past tense morpheme, since it surfaces as ***hit*** and not ****hitted*** The problem is that no explicit mechanism has been proposed for deciding that ***hit*** is an ambiguous string (that is, it corresponds to two different morphological feature sets), whereas, say Old Icelandic ***skip*** is a non-ambiguous form, unspecified for case. This lack of explicitness is unacceptable in a formal theory, though we find it in a variety of frameworks from Distributed Morphology to Lexical Functional Grammar.

While it is not the case that scholars who advocate broad use of vague representations deny the possibility of homophony, it is useful to demonstrate that homophony is present in cases more subtle than, say, English *well*, a string which corresponds to both a noun and an unrelated adverb, or *knight* and ***night.*** In this section, I present five distinct arguments that homophonous or ambiguous representations must be allowed in morphological theory. This evidence should serve to steer us further and further away from radical vagueness.

VI.1. First argument: Blocking of productive morphology

English has a productive marker of plurality in nouns: *rodent-s,* ***banana-S,linguist-s.*** Following early discussion of blocking in morphology (Aronoff 1976), the form ***feet*** can be said to block the derivation of **foots*. In other words the existence of a stored form FOOT-PLURAL prevents the concatenation of the independently listed FOOT and PLURAL lexical items.

A more interesting case is that of the 'irregular' noun ***sheep*** which could potentially be treated as ambiguously singular or plural (corresponding to two representations), or as unmarked with respect to number (i.e., vague in interpretation). However, in this case, there must be a lexical item '***sheep*** [PLURAL]' in order to block the productive process of plural formation from generating **sheeps*. Therefore, we can conclude that at least in some cases, forms which are

superficially ambiguous as to whether they are marked for a given category, are in fact ambiguous. Without a (possibly phonologically null) featural distinction, there is no mechanism available to block the productive morphology from generating ***sheeps***[3].

VI.2. Second argument: identical subsumption structures

Vague representations must be formulated in such a manner that the features they contain are compatible with all the environments in which the representation is used. I assume that listing disjunctive subsets of features associated with a single phonological representation is a notational variant of positing distinct, homophonous representations. Vague representations, therefore, must contain, some subset (possibly not a proper subset) of the set derived by the intersection of the environments in which they can be used[4]. Depending on the set of features used and depending on whether an attempt is made to make representations as 'economical' as possible, theories of morphology have appealed to a principle, such as the Elsewhere Principle, that specifies which representation is chosen if more than one is compatible with a given context. We consider now potentially problematic cases for such a theory.

Consider the paradigm in (10) which is typical of Old French masculine nouns. The example is the word 'wall'.

(10) An Old French problem for vague representations

	NOM.	**OBL.**
SING.	*murs*	*mur*
PLUR.	*mur*	*murs*

We see that the nominative singular and oblique plural are both *murs*, whereas the oblique singular and nominative plural are both ***mur***. Therefore, we have to recognize that the same phonological form corresponds to different feature specifications in these paradigms.

To show that this is not an isolated case, a further example of this situation can be supplied by Scots Gaelic (Calder 1923: 81, 101).

(11) A Scots Gaelic problem for vague representations

		OBL.
SING.		*àil*
PLUR.	*àil*	

The nominative singular and genitive plural of the word for 'brood' have a so-called 'broad vowel' form, whereas the genitive singular and nominative plural show a 'slender infection' *[sic]*.

We can now define the general case illustrated by these examples. Given a paradigm defined by contrasting pairs of features (or feature values) F and G, and A and B, the theory of vague specification runs into trouble if it is the case both (1) that the exponents of [F, B] and [G, A] are identical, say X, and (2) that the exponents of [F, A] and [G, B] are identical, say, Y:

(12) A hypothetical problem for vague representations

	A	B
F	*Y*	*X*
G	*X*	*Y*

The positions filled by X correspond to [F, B] and [G, A], which means they share no features. Their intersections for the relevant features is the null set. The positions filled by Y are [F, A] and [G, B] which also share no relevant features and thus, their intersection for the relevant features [A, B, F, G] is also the null set[5]. So the underspecified 'vague' representations for X and Y are identical, both are the empty set.' Therefore, there is no way for the morphology to choose among them without resorting to disjunctions like "Choose Y if either [F] and [A], or [G] and [B]". This just moves the responsibility for keeping track of the correct form from the lexicon to an ad ***hoc*** rule of lexical selection. Given the fact that X and Y can, in principle, be morphologically complex and show the effects of phonological neutralizations, it is unclear how such a rule could be formulated. It is also impossible to choose X or Y as more marked, that is more specific, using set intersection as the procedure to determine which features each item is specified for.

VI.3. Third argument: neutralization in morphophonologically definable contexts

An example from English suggests further that the practice of collapsing homophones that share a 'significant' portion of their featural makeup is excessively superficial. Many dialects of English do not distinguish phonologically the possessive and non-possessive forms of regular plural nouns: ***the girls*** *left* and ***the girls'*** *mother left* both contain the phonological string [grlz]. A theory that demands the collapse of subject and object forms of English nouns, would seem to require the collapse of the possessive and non-possessive plurals. Furthermore, it seems that we would also need to collapse the possessive singular form ***girl's*** which is homophonous with the other two. This leaves the form ***girl*** as the most marked member of the paradigm, requiring

explicit encoding of non-plurality and non-possessiveness —a conclusion we assume is odious to believers in markedness. Of course, one could conclude that this is exactly the point of markedness, and take such a case as evidence of the marked nature of bare noun stems like *girl* in English. Note, however, that this collapse will not be posited for cases like *children's*. The analysis fails to capture the effects of accidental homophony.

VI.4. Fourth argument: Lexical splits

Toivonen (2000) demonstrates that the long-standing problem of the distribution of the Finnish possessive suffixes can only be solved by recognizing that the suffixes represent, in fact, pairs of homophonous forms. For the first and second person suffixes the argument is based primarily on principles internal to Lexical- Functional Grammar, the framework Toivonen adopts, though the conclusions may be compatible with other theories. In brief, in the absence of an independent pronoun the first and second person suffixes have a PRED feature and thus are pronominal suffixes, whereas when an independent pronoun is present the suffixes must lack a PRED feature. This avoids a PRED clash, since the pronouns have their own PRED feature. In such a case, the suffixes are merely agreement markers. For these persons, the representation of the agreement marker is contained in the representation of the form which has a PRED.

The third person suffix *-nsa/nsä* provides a further compelling demonstration. Toivonen shows that in the third person, in contrast to the first and second, the representations of the agreement marker and the form used without an independent pronoun are in fact distinct —neither representation subsumes the other. The agreement marker agrees only with a third person, HUMAN pronoun:

(13) (a) *Pekka nakee hänen ystävä-nsä.*
P. sees his/her friend-3Px
'Pekka sees his/her friend'

(b) **Pekka nakee hänen ystävän.*
P. sees his/her friend. ACC
'Pekka sees hisher friend'

(c) *Pekka nakee pojan ystävän.*
P. sees boy.GEN friend.ACC
'P. sees the boy's friend'

(d) **Pekka nakee pojan ystävä-nsä.*
P. sees boy.GEN friend-3Px

(e) ***Mina annan koira-lle sen ruokaa.***
I give dog.ALL it.GEN food
'I give the dog its food'

(f) ***Mina annan koira-lle sen ruokaa-nsa.***
I give dog.ALL it.GEN food-3Px

In (13a), the possessive suffix *-nsa/nsä* agrees with the third person, HUMAN pronoun *hänen*. In (b), the sentence is ungrammatical without the suffix. In (c), we see that no suffix is present when there is a full lexical NP possessor. In (d) the suffix is meant to agree with the full lexical NP possessor, which is obviously not a pronominal form-this is ungrammatical. In (e) there is a pronoun, but it is ***sen,*** which cannot refer to humans, so the sentence is grammatical without the suffix; and ***(f)*** is ungrammatical with the suffix, because of the non-human pronoun ***sen.***

In contrast to such cases, the suffixed pronominal form, which must be anaphoric with the subject, is not restricted to reference with humans. It can corefer with any third person even those that are non-human and non-animate, and its antecedent need not be a pronoun:

(14) (a) ***Han nakee ystävä-nsä.***
He sees friend-Vx
'He, sees his_i friend'

(b) ***Poika nakee ystava-nsa***
boy sees friend-Vx
'The boy_i sees his, friend'

(c) ***Se heiluttaa häntää-nsä.***
it wiggles tail-3Px
'It_i wiggles its, tail'

Thus, the morphological features of the homophonous 3rd person possessive suffixes are distinct. Slightly simplifying Toivonen's discussion, we can represent the features of the agreement marker as in (15):

(15) Features of agreement marker *-nsa/nsä*

HUMAN
3rd
PRONOUN AGREEMENT

And we can represent the features of the pronominal suffix as in (16):

(16) Features of pronominal suffix *-nsa/nsä*

PRED
SUBJECT ANTECEDENT
3rd

The surface form *-nsa/nsä* is thus ambiguous—there are two homophonous forms. It is possible to list, say, a disjunctive statement of where the putative 'vocabulary item' *-nsa/nsä* is inserted. However, this is equivalent to listing two separate items.

VI.5. Fifth argument: Evidence for zero derivation

It is well known that English has at least two verbsfly. One appears in sentences like That bird *flies* out of the barn whenever rhe cat comes in. Another appears in sentences like Thepitcher *flies* out to right *field* every time he bats. The past tense of these verbs differ. The first corresponds to irregularflew, whereas the baseball term can only have the past tense formflied. Irregularness is a feature of roots, and the first verb is based on a verbal root which is linked to the irregular *flew*. The baseball verb, however, is (synchronically) productively derived from the nounfly *(ball)* and thus is not a stored form. For our purposes, what is relevant is the obvious point that despite their homophony and intuitive similarity in meaning, the two verbsfly must correspond to distinct representations. Therefore, the phonological formfly is ambiguous, not vague.

VII. SOLVING THE LINGUIST'S PROBLEM—L2 DATA TO THE RESCUE

It is now clear that there is some homophony, so radical vagueness is untenable. As mentioned above, we do not have any explicit algorithm for determining exactly how much homophony there is, yet it is intuitively wrong that English has distinct DUAL and PLURAL forms of every noun. We cannot come to a complete solution to the determination of how much homophony each language has and what principles determine this, but it seems that languages lie somewhere between the two logical extremes in the mapping between phonological strings and lexical entries. (See Reiss (2000) for discussion of explicit learning algorithms that differ in the amount of homophony that they lead to). In this section, I will just show how L2 data can bear on the issue of homophony in particular cases by helping us decide if a given phonological string

corresponds to one or more lexical entries.

Since, for example, an Old Icelandic grammar with two distinct (homophonous) words pronounced ***skip*** will be extensionally equivalent in terms of the phonological strings generated to a grammar that contains a single vague item pronounced ***skip,*** intemal evidence can never be used to choose among the two (by definition). There is the possibility, however, of appealing to external evidence from SECOND LANGUAGE ACQUISITION (SLA).

The errors that SLA leamers make may reflect aspects of the L1. Consider the following (impressionistic) observations: speakers of a language like Hungarian, which does not distinguish gender in third person pronouns make many errors in using English *he/she*, whereas English speakers do not appear to have a problem leaming not to be able to distinguish the genders. If Hungarian o corresponded to two separate representations, one [3 SG MASC] and another [3 SG FEM], we might expect the mapping to the English system to be easier than it apparently is. Similarly, English speakers leaming Marshallese have a hard time leaming to make the DUAL/PLURAL contrast, so this may indicate that this distinction has been collapsed in English grammars. In other words, we can reject, on the basis of external evidence, the idea (Radical Ambiguity) that no collapse of initial full specification occurs. Hungarian speakers do not have two (or more) third person pronouns, and English speakers do not have a covert DUAL. So not every distinction allowed by UG remains encoded in every grammar.

L2 data also provides further evidence some covert distinctions do exist. Do English speakers have a problem leaming distinctions like the French *tu/vous* contrast? My intuition is that they do not. If correct, this can be taken as evidence that distinctions that are made anywhere in the language, such as SINGULAR vs. PLURAL, are maintained in all relevant representations. In other words, the evidence from English speakers leaming French or Spanish suggests that English has at least two pronouns that are pronounced ***you.*** The question of whether there are more than two is discussed further in Reiss (2000). This conclusion may be contrary to the intuítions of many morphologists, but it is based on an explicit form of reasoning and can be empirically tested. Therefore, it should be taken seriously.

Another example, is perhaps easier to swallow. In Italian, the string *sono* corresponds to both ***I am*** and ***they are.*** Here, I think, the intuitions of linguists are in agreement that this case is one of accidental homophony, but it is not clear that this intuition has ever been justified. However, consistent with the intuition is the fact that Italian leamers of English will not confuse ***am*** and ***are,*** as we might expect them to if there were a one-to-two mapping from Italian lexical items to English ones. So, *sono* is ambiguous—it corresponds to two homophonous items in the Italian lexicon.

The general point is just this: one-to-one and many-to-one mappings fromL1 to L2 are easier than one-to-many mappings. In other words, it is easier to map to the correct form if there is only one output choice than ifthere is more than one. The error pattems of L2 leamers can give us insight into the nature of the mappings, and thus into the nature of theL1 grammar.

Only by taking issues of L2 acquisition into account, can we start to solve the *Linguist's Problem* of the indeterminacy of the data and arrive at a theory of how the *Human's Problem* is solved. Such simple arguments suggest that theoretical linguists would do well to consider L2 data more carefully.

NOTES:

1. It is not even worth arguing about this assumption of the nativist hypothesis—nobody denies that there is something innate that makes us able to learn languages and does not allow chickens to do so upon comparable exposure.

2. Note the same sloppy terminology is rampant in syntax. We considera sentence to have a hierarchical structure, as well as alinearorder. Therefore, sentences cannot be structurally ambiguous, only strings can be structurally ambiguous.

3. Of course, one could stipulate that the plural morpheme is linked to a list of roots which it cannot attach to. This kind of 'negative subcategorization' is typically not considered in generative analyses and we will not pursue the idea here.

4. Given the likelihood that the relevant morphosyntactic features are organized somewhat, and not elements of an (unstructured) set, it would be more appropriate to talk about subsumption relations than mere subset relations. However. I forego this (ultimately necessary) distinction here, for the sake of expository convenience.

5. Obviously, the relevant lexical items, X and Y are specified, in fact they are identically specified, for further features not listed here.

REFERENCES

Aronoff, M. **(1976).** ***Word Formation in Generative Grammar. Linguistic Znquiry Monograph no. 1.*** Cambridge, MA: MIT Press.

Bresnan, J. **(1997).** Explaining morphosyntactic competition. To appear in ***Handbook of Contemporary Syntactic Theory,*** ed. by Mark Baltin and Chris Collins. Oxford: Blackwell.

Calder, G. **(1972).** A ***Gaelic grammar***. Glasgow: Gairm Publications.

Chomsky, N. **(1986).** ***Knowledge of Language.*** Westport: Praeger.

Dalrymple, M. & R. Kaplan **(2001).** *Feature structures*. Language.

Grimshaw, J. **(1997).** Projection, heads, and optimality. ***Linguistic Znquiry*** **28,373-422.**

Hale, M. & C. Reiss **(1998).** Formal and empirical arguments concerning phonological acquisition.

Linguistic Inquiiy 29, 656-683.

Kager, R. (1999). ***Optimality Theory.*** Cambridge: Cambridge University Press.

McCarthy, J. (1999). Sympathy, Cumulativity, and the Duke of York Gambit. ***ROA-315.***

Prince, A. & P. Smolensky (1993). ***Optimality theoiy: constraint interaction in generative grammar.*** Technical Report RUCCS, Rutgers University, New Brunswick, N.J.

Pylyshyn, Z. (1984). ***Consciousness and the Computational Brain.*** Cambridge, MA: MIT Press.

Quine, W.V. (1972). Methodological reflections on current linguistic theory. In G. Harman and D. Davidson (eds.) ***Semantics of Natural Language.*** New York: Humanities Press.

Reiss, C. (2000). Homophony. Ms. Concordia University.

Reiss, C. (1994). The outcome of /lr/ and /nr/ in Old Icelandic. ***Linguistic Review*** 11,329-350. Warning: This article is full of important typos and misprints! Ask me for a corrected copy if it interests you.

Tesar, B. & P. Smolensky (1998). Leamability in Optimality Theory. ***Linguistic Inquiry 29,*** 229-68.

Toivonen, I. (2000). The Morphosyntax of Finnish Possession. ***NLLT*** 18: 579-609.

