Separating Input from Intake: Acquiring Noun Classes in Tsez

Annie Gagliardi & Jeffrey Lidz

Department of Linguistics
University of Maryland, College Park

1401 Marie Mount Hall
University of Maryland
College Park, MD 20742-7505
Tel: +1-301-405-8220
Fax: +1-301-405-7104

acg39@umd.edu
jlidz@umd.edu

Abstract

This paper examines the acquisition of noun classes in Tsez, looking in particular at the role of noun internal distributional cues to class. We analyze a new corpus of child directed Tsez speech, determining the proportion of nouns that children hear with this predictive information and how often this is heard in conjunction with overt noun class agreement information. Additionally we report on asymmetries in the classification of nouns with predictive features in the corpus and by children and adults in an elicited production experiment. We show that children use noun internal distributional information as a cue to noun class out of proportion with its reliability. Instead, children are biased to use phonological over semantic information, despite a statistical asymmetry in the other direction.

1 Introduction

Perhaps one of the most rehearsed stories in linguistics concerns children's uncanny ability to acquire language. While all children acquire the structure of their native language in a mere five (or so) years, with little apparent effort or confusion, language scientists fare considerably worse in identifying that structure. Teams of linguists have been studying linguistic structure for millennia and nonetheless continue to discover new generalizations and struggle to find the appropriate representations for capturing them. This story, or so it goes, reveals the special talent that human children (as opposed to human adults, chimps, rats or professional linguists) have for acquiring language and suggests that children bring to the language learning task an innate stock of implicit representations and analytic tools that allows them to see through the vagaries of linguistic distribution in order to home in on the appropriate representation of the language in their environment (Chomsky 1959, 1965; Gold 1967, Pinker 1979, Crain 1991, Jackendoff 2002; among many others). The study of children's language learning in this context largely amounts to an investigation of how children project beyond what could reasonably be inferred from their experience.

This story is typically offered in response to learning theories based solely on distributional analysis (e.g., Harris 1951, Rumelhart & McClelland 1986, Elman et al 1996), in which the learner builds the structure of the language piecemeal by first using the distribution of phones to find the significant phonological generalizations, then analyzes these to discover the morphological structure and so on, up to syntax, semantics and pragmatics. In recent years, however, the role of distributional analysis has taken on renewed interest as the computational tools for conducting such analyses have become more sophisticated and potentially offer a reconsideration of arguments for the insufficiency of distributional analysis as a model for language acquisition. The study of language acquisition, from this perspective, amounts to rigorous computational analysis of what is in principle inferrable from linguistic experience (in the absence of explicit constraints on the character of linguistic structure) and attempts to bring these into alignment with how children actually develop (e.g., Lewis & Elman 2001, Ambridge et al 2009)

In the current paper, we consider a case where children seem to fare considerably worse than linguists, drawing different conclusions from the statistical information than a purely distributional learner would. Whereas the linguist armed with some simple tools of distributional analysis can identify the noun classes of a language in a relatively brief time, children apparently struggle with this into the school years (MacWhinney 1978, Karmiloff-Smith 1979, Mills 1986). The acquisition of noun classes ought to be trivially easy. Each noun occurs in agreeing contexts some proportion of the time and the agreeing element consistently exhibits the appropriate agreement. We argue that the inferiority of children's performance in noun classification to that of linguists and computational models is as informative as the reverse about the tools that learners bring to the task of acquiring a language. In particular, we argue that such cases allow us to separate the role of the *input*, or the actual information present in the linguistic environment, from the role of the *intake*, or the information from the input made available by the learning mechanism, in language acquisition, giving us some insight into the particular distributional analyses that children are prepared to engage in.

As just noted, learning noun classes should be easy. There are two types of information that can be used to characterize noun classes. First there is what we will call noun *external* distributional information: agreement information in syntactic context that reflects the class of the noun triggering agreement. Second there is noun *internal* distributional information: semantic or phonological similarities among the nouns in a given class. Until we determine whether or not children make use of this information as a cue to noun class, we will conservatively call these noun external and noun internal properties 'information', and not 'cues'. By looking at noun external distributional information a trained linguist could sit down with a language and quickly determine (1) whether the language in question had noun classes (2) how many classes there were and (3) which class each noun used with agreement went into. With just a little more work the linguist could also determine similarities among the nouns in each class and use these with varying degrees of success to predict the class of nouns not previously seen with agreement (see Corbett 1991 for review). These two kinds of information: the highly regular noun *external* distributional properties (syntactic context) and the probabilistic noun *internal* distributional properties (similarities among properties of nouns within a class that vary in their reliability) are presumably available in abundance to the learner. If they weren't, the language in question wouldn't have a noun class system.

With both highly regular and probabilistic information in principle available to the learner, we can ask what information the learner actually makes use of when going through the same steps of discovering noun classes and the properties that correlate with them. That is, what of the available information in the input is used as a cue in the intake. While it may look like there is ample evidence for the existence and structure of the noun classes in the input, what portion of this evidence is actually used depends on more than just what information is available - it also depends on how this input is filtered by the learning mechanism when it is taken in (Lidz & Pearl 2007). This is an area where we must distinguish between the input and the intake. Because children acquiring language can get so far from seemingly so little information in other cases, it is an intriguing puzzle to study what they do when a seeming overabundance of information is available. Does the learner make use of all available information? Is all the information that appears available to the researcher really available to the learner? If not, what sort of intake mechanism is responsible for the filtering of the input and why?

In this paper, we do not directly investigate how the learner initially discovers noun classes, but instead look at a learner with a developing system of noun classes. By looking at how this developing system differs from the adult system we can glean information about (1) how the learner thinks nouns are organized into classes and (2) what of the available information the learner must have used to arrive at this state. These two pieces of evidence allow us to draw inferences regarding discovery of noun classes earlier in development. In particular we consider two hypotheses regarding the acquisition of noun classes: the External Only Hypothesis (EOH) which suggests that children only use noun external distributional information to determine the existence and composition of noun classes and the Hybrid Hypothesis (HH), which suggests that both noun external and noun internal distributional information play a key role in noun class acquisition. Not only do children appear to use both types of information, supporting the HH, they appear to use noun internal distributional information out of proportion with its statistical reliability. Despite a statistical asymmetry in the input where semantic information is a more reliable predictor of class than phonological information, the intake appears to be biased towards phonological over semantic information.

Section 2 details what noun internal and external information actually look like in Tsez, a Nakh-Dagestanian language we will use to investigate the acquisition of noun classes. Section 3 lays out two hypotheses relating noun external and noun internal distributional information to the acquisition of noun classes: the External Only Hypothesis, positing that children only use noun external information in the acquisition of noun classes, and the Hybrid Hypothesis, positing that children use both noun internal and noun external information. Section 3 also gives an overview of related work. Section 4 presents a new corpus of child directed Tsez, and an analysis of this corpus that reveals what noun internal and noun external information is available to the learner and crucially determines the statistical reliability of noun internal information. Section 5 contains the key observation of the paper: where behavioral experiments with adult and child Tsez speakers reveal an asymmetry between the sensitivity of children to noun internal information and the behavior predicted by the reliability of this information. Section 6 shows how the experimental findings support the Hybrid Hypothesis, and relates them back to the input/intake distinction. Finally, several hypotheses accounting for the existence of this distinction are put forward .

2 An overview of noun classes in Tsez

Natural languages all over the world employ noun classification systems. These systems can generally be divided into two types: noun class (or gender[1]) systems and classifier systems. In noun class systems, the class of a given noun can influence the form of items in the entire sentence, whereas in classifier systems the influence of the class of a noun is limited to the noun phrase. This paper focuses on noun class systems, but similar arguments could be applied to the acquisition of classifier systems. Noun classes can be characterized in two ways: using the noun *external* distributional properties such as the agreement paradigm or syntactic behavior that defines the class and using noun *internal* distributional properties, the characteristics of the nouns that make up each class. As mentioned above these two types of information could be used in noun class acquisition[2].

2.1 Noun External Distributional Properties

[1] Corbett (1991) refers to all noun classification systems as grammatical gender, whether the system makes use of natural gender or not. We agree that this is the correct, as both systems have the same sorts of grammatical reflexes and their acquisition should be governed by the same mechanism. In our experience, a significant degree of confusion arises when noun classification systems that make use of natural gender (but differ from purely gender based systems such as the English pronominal paradigm) are called ‘genders’. Therefore in this paper we will use the term noun class, as it suggests no primacy of certain correlating features over others.

[2] Certain types of verb classes might be superficially characterized in a similar way - members of a class both share external properties such as the tense morphology they exhibit, and internal properties such as phonological form or even meaning, and so in some cases it might be appropriate to investigate their acquisition and representation in a parallel fashion.

Noun classes are defined as groups of nouns that pattern the same way with respect to agreement. Languages differ as to where this agreement is seen (Corbett 1991). Some languages are limited to DP internal agreement[3], appearing on pronouns, possessives, numerals, determiners and adjectives. Other languages also allow agreement external to the DP, on verbs, adverbs, adpositions, complementizers and even other nouns. Languages vary greatly in terms of how many environments agreement appears in. They also vary in terms of the number of classes, some with as few as two (Spanish, French) and others with as may as 20 (Fula) (Corbett 1991).

For a more concrete example, consider Tsez, a Nakh-Dagestanian language spoken by about 6,000 speakers in the Northeast Caucasus[4]. Tsez has four noun classes in the singular which collapse to two in the plural. The noun external distributional information characterizing these classes is prefixal agreement on vowel initial[5] verbs, adjectives and adverbs, as shown in Table 1.

Table 1: Tsez Singular Noun Class Agreement

Class 1	**Class 2**	**Class 3**	**Class 4**
Ø-igu uži	**j**-igu kid	**b**-igu k'et'u	**r**-igu čorpa
I-good boy(I)	II-good girl(II)	III-good cat(III)	IV-good soup(IV)
good boy	*good girl*	*good cat*	*good soup*

[3] Again, contrasting with classifiers, which appear to be restricted to the NP

[4] According to the 2002 census, there are about 15 thousand Tsez speakers, but the real number estimated by researchers is around six thousand (Bokarev 1967, Comrie et al. 1998; Comrie and Polinsky 1998; Polinsky 2000).

[5] A small proportion of verbs, adjectives and adverbs are vowel initial but do not take overt agreement. An interesting observation to make would be whether children overgeneralize agreement to these exceptions.

Thus the agreement prefix for class 1 is the null prefix, for class 2 it is [j], for class 3 [b] and class 4 [r]. The same set of prefixes are used on verbs, adjectives and adverbs. Plural agreement prefixes and some forms of both personal and demonstrative pronouns also vary by noun class, but there is considerable syncretism in these paradigms, making them less reliable markers of class (Tables 2-4).

Table 2: Tsez Plural Noun Class Agreement

Class 1	Class 2	Class 3	Class 4
b-igu uži-bi	**r**-igu kid-bi	**r**-igu k'et'u-bi	**r**-igu čorpa-bi
I-good boy(I)-abs.pl	II-good girl(II)-abs.pl	III-good cat(III)-abs.pl	IV-good soup(IV)-abs.pl
good boys	*good girls*	*good cats*	*good soups*

Table 3: Tsez Personal Pronouns[6]

<table>
<tr><th></th><th></th><th>Class 1 (singular)</th><th>Class 2-4 (singular)</th><th>Class 1 (plural)</th><th>Class 2-4 (plural)</th></tr>
<tr><td rowspan="3">1st Person</td><td>Absolutive</td><td colspan="2">di</td><td>eli</td><td>ela</td></tr>
<tr><td>Oblique</td><td colspan="2">dā-</td><td>elu-</td><td>ela-</td></tr>
<tr><td>Genitive</td><td colspan="2">dej</td><td colspan="2">eli, eliz</td></tr>
<tr><td rowspan="3">2nd Person</td><td>Absolutive</td><td colspan="2">mi</td><td>meži</td><td>meža</td></tr>
<tr><td>Oblique</td><td colspan="2">debe-, dow-</td><td>mežu-</td><td>meža-</td></tr>
<tr><td>Genitive</td><td colspan="2">debi</td><td colspan="2">meži, mežiz</td></tr>
</table>

[6] Tsez only has personal pronouns for 1st and 2nd person. Demonstrative pronouns are used as 3rd person pronouns. Effectively the personal pronouns are only used with classes 1 and 2, as they will generally have human referents. However, in stories or other contexts where non human nouns might be referred to in the 1st or 2nd person, they require the same pronouns as class 2.

Table 3: Tsez Demonstrative Pronouns

		Class 1 (singular)	Class 2-4 (singular)	Class 1 (plural)	Class 2-4 (plural)
Proximal	Absolutive	-da	-du	ziri	
	Oblique	-si	-ła-, -ł	-zi	-za
Distal	Absolutive	že		žedi	
	Oblique	nesi	neło, neł	žedu	žeda

In any language with a noun class system, seeing an agreement marker for a given class used in conjunction with a noun is a signal that the noun is in the class corresponding to the agreement marker. In Tsez, only the singular noun class agreement unambiguously signals the class of any noun. For a linguist setting out to determine what class each noun is in, looking at the singular agreement that goes along with each noun is enough to discover that classes exist, to determine the number of classes in the language and to determine the class of each noun. It could be that this is also how a child accomplishes both tasks. While the syncretism evident in the plural and pronominal paradigms might make this task more difficult for the child, this will be true whether the child is only using noun external distributional information to acquire noun classes or not. Because only singular agreement provides reliable evidence for the existence of 4 classes, we restrict our attention to the singular agreement marking for the remainder of the paper.

2.2 Noun Internal Distributional Properties

If noun internal distributional information is important for the acquisition of noun classes, it is imperative to determine whether or not languages have, for each class, some feature or set of features characteristic of the nouns in that class. The results of many typological surveys is resoundingly positive: every noun class system appears to have some regularity in the way at least a subset of nouns are classified (Corbett 1991), and that could be enough to aid the learner. For the acquisition researcher investigating whether or not these regularities are employed in noun class acquisition, it does not matter whether there is a set of rules that can classify all nouns based on noun internal distributional information, or merely a subset. If some noun internal information correlates with class, that is enough to launch an investigation to determine whether or not the child makes use of this information during acquisition. Below we will look at the noun internal distributional information that characterizes Tsez noun classes.

Plaster et al investigated the cues that characterize nouns in Tsez, and found that rules could classify nearly 70% of nouns in the dictionary. A summary of the classes based on traditional descriptions of the language (Comrie and Polinsky 1999) is found in Table 5:

Table 5: Summary of Tsez Noun Classes

Class 1	Class 2	Class 3	Class 4
all male humans	all female humans	all other animates	many other things
only male humans	many other things	many other things	
~13% of nouns	~12% of nouns	~41% of nouns	~34% of nouns

percentages reflect the percentage of the nouns in class in the dictionary (Khalilov 1999)

Class 1 is perhaps the most unusual class, consisting of all male humans and only male humans. This means that the assignment of new words to Class 1 is more restricted than any other class. Not reflected in percentages are nouns that can also refer to female humans in the right context (such as *teacher*), which are then used with Class 2 agreement, as all female humans belong in Class 2. Unlike Class 1 however, the majority of the class is comprised of inanimate or abstract nouns. Class 3 is the largest class and, while it contains all animate, non human entities, it also contains a wide variety of inanimate and abstract nouns. Class 4 contains many inanimates and abstracts, including a morphologically derived set of abstract nouns ending in the suffix [-ɬi]. While these generalizations can be used to classify roughly 25% of Tsez nouns, they do not approach exhaustive classification.

Plaster et al took the set of nouns from a Tsez dictionary (Khalilov 1999), and tagged them for possibly predictive features. These features included semantic features such as animacy and various physical and functional properties, phonological features such as first and last segments and morphemes, number of syllables and formal features such as the declension class. The result was a feature vector for every noun that included values for every possible feature for a given noun. The set of feature vectors was the input to a supervised learning algorithm, Quinlan's C4.5 implementation of a decision tree algorithm (Quinlan 1993). The output of such an algorithm is a set of decision rules, dependent on the presence or absence of a certain feature on a noun, determining classification of the noun or the next decision to be made. For example, since the feature *male human* is a very reliable feature that can be used to reliably classify a large number of words, the first rule in the decision tree assigns all nouns with the feature *male human* to Class 1. Nouns without this feature are then subject to the next rule, and so on, until all nouns have been classified.

By using the sorts of features described above in such an algorithm, Plaster et al were able to accurately classify about 70% of Tsez nouns. Semantic features, both those referencing properties like animacy and humanness and those referencing physical properties like stone or container were found to be more predictive than formal properties such as certain derivational suffixes and the first segment of the noun. This number looks promising, considering the large degree of arbitrariness that the Tsez system at first appeared to have. While Plaster et al see this as only a good first pass, and endeavor to better characterize the classification of the remaining 30% of nouns, the fact that several features can be reliably used to predict noun class is as much as we need to move forward investigating their role in the acquisition of noun classes.

3 The Role of Noun External and Noun Internal Distributional Properties

Now that we have outlined the two types of information that are in principle available in the input to the learner of Tsez we can hypothesize what information makes up the intake, and how how this information may be used. There are two senses in which they could be used: by adults to both represent their noun class systems and to classify novel nouns, and by children to acquire the system of classes and classify nouns as they learn them.

In the discussion that follows, we will assume that in the adult representation of noun classes, class is stored along with the lexical entry of a given noun and is accessed every time a noun is processed or produced, but not repeatedly recomputed based on internal or external information. We assume that children are acquiring the same sort of system that adults have.

3.1 Adult Representation and Classification of Nouns

It is evident from adult speakers' use of their native language that they can use noun external distributional properties when processing sentences, and presumably this information is diagnostic of the class of novel nouns as well. That is, if an adult speaker hears a word used in the syntactic context characteristic of a given class, he or she will know that the novel word belongs to that class. This information is highly regular in the language as it provides the characteristic definition of the class, and is thus presumably a very reliable cue to the class of a novel word.

Evidence from borrowings and previous research (Tucker et al 1977, Corbett 1991, Polinsky & Jackson 1999) shows that adults can also use noun internal distributional information to classify novel nouns in the absence of the more reliable syntactic information. Novel nouns that have noun internal properties in common with a group of nouns in a given class are likely to be put into that class. Exactly how this works though, is not clear. Do speakers have a set of classification rules associated with predictive noun internal properties (e.g. If a noun denotes a female human, then classify it as Class 2)? Or do the predictive noun internal properties inflate the probability that a noun would be in each class in favor of the class that that property predicts (e.g. within the existing lexicon it is 100% probable that if a noun denotes a female human it is in class 2, therefore novel nouns denoting female humans have a high probability of ending up in class 2)?

At this point it is relevant to relate noun class systems to other lexical subclass systems that also appear to share both external grammatical properties (e.g. past tense inflection) and internal properties (e.g. phonological form). For example, consider the subclass of English irregular verbs *ring*, *sing*, *drink*, *sink*. All of these verbs inflect for past tense via ablaut (*ring-rang*) and also share the [iŋ[+velar]] form. However, neither the existence of the *i-a* ablaut nor the [iŋ[+velar]] form is predictive of the other (e.g. *spit-spat*, *link-*lank*). Analyses posit that classes like these are represented as a class of exceptions to a regular rule (Pinker 1991), multiple rules acting over a small classes of words that tend to have phonological similarities (Halle & Mohanan 1985; Yang 2002) or are part of a system where grammatical reflexes apply probabilistically to classes of words with varying levels of similarities (Hay & Baayen 2005). It may be tempting to try to align the representation of noun classes to one of these analyses. However, differences in the way noun classes and this set of verb classes work mean that none of these analyses is appropriate for noun classes. We will expand on this observation in section 6, and also suggest that our analysis of noun classification may be applicable to irregular verb classes.

Returning to noun classes, we don't know at this point whether predictive information is used to determine which classification rule to apply, or to calculate the probability that a noun will fall into a given class. Even without fully specifying the exact nature of the rule system or the probability system, these two alternatives appear to make distinct predictions for the classification of nonce words. A rule system predicts that if there is a rule based on a certain feature, and this feature is observed on a novel word, it should be consistently classified according to this rule. A probability system predicts that if nouns with a certain feature have some probability distribution across classes, if this feature is observed on a novel word, the probability that the novel word is in a given class will be proportional to the probability of (1) the probability distribution of nouns with this cue (2) the prior probability of each class and (3) the probabilities associated with any other predictive features this noun contains. By precisely specifying what this probability is we can precisely model the classification of novel words. This modeling falls outside the scope of the current work and is addressed in forthcoming work (Gagliardi, Feldman & Lidz 2012). What is important for this paper is that predictive information will work either deterministically, as in a ruled based system, or probabilistically, as in a probabilistic system.

The question of whether predictive information is used for determining rules or calculating probabilities also becomes relevant when looking at the classification of novel words without identifiable predictive information. A rule based system will have a default classification rule for such nouns, whereas a probability based system will classify these nouns based on both the prior probabilities of each class and the probabilities of each class associated with not having certain predictive features.

3.2 Acquisition of Noun Classes

No matter the precise way in which noun internal distributional information works, in order to acquire a noun class system, to arrive at the system that adults exhibit - where noun external information is accurately produced and interpreted and speakers are sensitive to noun internal cues that correlate with class - children must at some point pay attention to both noun internal and noun external distributional properties. How and when they do this is the focus of the current paper. Recall that in order to acquire noun classes the learner must (1) notice that the language has noun classes (2) determine how many classes there are and (3) determine which nouns go in which classes. Below we outline two hypotheses regarding how these three steps may occur, as well as the predictions that each of these hypotheses makes for later behavior, allowing us to infer which of the hypotheses regarding earlier steps is consistent with later behavioral data.

There are two routes a child could take to acquire a noun class system that is actively characterized by both noun internal and noun external distributional information. The child could simply use noun external distributional information in the beginning to discover classes and classify nouns as they were encountered with telltale agreement. Such a system is similar to that outlined in Pinker (1984). Pinker proposes that a child learns morphological paradigms by filling in each cell with affixes encountered in the input[7]. When two affixes compete for entry in the same cell, the cell splits and two classes are formed. That is, a child might be filling in an agreement paradigm, and would discover another class when two different agreement

[7] This is a general paradigm building model proposed in Pinker (1984). It is distinct from the words and rules model developed later (Pinker 1991) and referenced above when discussing differences between the problem of representing noun classes and English irregular verbs.

morphemes competed for the same ‘verb agreement’ slot in the paradigm. Such a system does not rely on noun internal distributional information, only noun external distributional information such as agreement. Instead, for children to acquire adult-like sensitivity to noun internal distributional properties, they would have to keep track of this information after the noun class system had been acquired. Once the lexicon has sufficient content the learner could generalize over items in each class to extract the noun internal distributional information, that is, the statistical regularities describing the nouns in each class. We will call this hypothesis the *External Only Hypothesis* (EOH).

A second hypothesis is that the child first uses only noun internal distributional information, grouping nouns together by their featural content, and at a second stage combines these many small groups of nouns to form classes, by noting the coocurrence of these subclasses of nouns with class dependent noun external distributional information. At a certain stage, they would be able to use the external rather than the internal distributional information to characterize a class. Such a process was suggested by Braine (1987) after observing that learners of artificial languages with lexical classes required both distributional information external to the items in each class and regularities internal to the items in a class, in order to discover the class system. Braine proposed a two step process wherein a learner first uses the internal information and later uses the noun external information. We will call this approach, the early use of both noun external and noun internal properties the *Hybrid Hypothesis* (HH).

These two hypotheses make different predictions regarding the differences between input and intake in noun class acquisition. The EOH predicts that children may not have a good command of the noun internal distributional properties characterizing nouns in a given class early in development, but that when they do acquire this sensitivity it should closely parallel that of adults. As the noun internal distributional properties would be calculated after the lexicon is well established, characteristics of both form and meaning should be equally well represented in the learner's achieved distributional sensitivity. That is, early in the development the intake will differ from the input, in that noun internal properties may not form a part of noun class representations, yet when these properties are incorporated they will be drawn from a mature lexicon, and will thus closely match the noun internal distributional properties attended to by adults. The HH on the other hand predicts that children should be sensitive to noun internal distributional properties from the earliest point at which they exhibit any knowledge of noun classes. As the lexicon is still being formed at this early stage, it is possible that the statistical regularities extracted early on will not reflect the actual regularities present in the input, and presumably used by the adult lexicon, but instead a version of these regularities filtered by the early intake mechanism.

3.3 Previous Research on the Acquisition of Noun Classes

Previous research on the acquisition of noun classes has shown that children acquiring noun class languages are sensitive to both noun external and noun internal distributional information, offering tentative support for the HH. Work in French (Karmiloff-Smith 1979), Spanish (Perez-

Pereira 1991), German (MacWhinney 1978; Mills 1985, 1986) and Russian (Rodina 2009) consistently shows that children are able to make use of noun internal distributional information in the classification of novel nouns. Moreover, younger children in particular prefer to use morphophonological information rather than semantic information, despite the fact that the semantic information in some cases is a more reliable predictor of class. Children also make use of noun external distributional information, though young children appear less able to do so.

Both the early reliance on noun internal distributional information and the fact that this reliance does not always align with the statistical reliability of the information as can be measured in the input point towards the HH. Unfortunately, this work does not directly address the questions posed by the hypothesis outlined above, as there are no direct comparisons with adult speakers and no information about what children or adults do when nouns are presented in the absence of either noun internal or noun external distributional information.

By examining the acquisition and representation of noun classes in Tsez, we will directly investigate (1) whether the EOH or the HH appears to be supported by the data and (2) whether noun internal information is employed in a rule or probability based system.

4 Information Available to the Tsez Acquiring Child: A Corpus Experiment

Above we discussed the two types of information characterizing noun classes in Tsez, and two hypotheses regarding the way in which this information could be used by a learner. Differences between the input as we can measure it and the intake, as can be inferred from behavioral data will help to differentiate between these hypotheses. In order to determine what of the input is used, we first have to characterize what exactly the input to a Tsez learner is. A limitation of the prior work on Tsez is that it is based solely on the distribution of words in the dictionary. Since learners are likely not exposed to the entire dictionary, we do not yet know what internal features of nouns are predictive of noun class in speech to children (and if these are different from the dictionary distributions), how often they hear nouns with these features, how often they are exposed to noun external distributional information and how often they hear these two types of information together. To address this issue, we created a corpus of child-directed speech in Tsez so that we could rigorously examine how much of this information is available in the input that learners actually receive. Once we have characterized what information the learner is exposed to, we can investigate hypotheses about how this information is used.

4.1 The Corpus

Over a period of 1 month, 10 hours of child directed speech were recorded during normal daily interactions between a mother, aunt and older sister of two 20-month-old Tsez acquiring children in Shamkhal, Dagestan. Roughly 6 hours of these recordings were transcribed with the assistance of two native speaker members of the family, familiar with the situations going on when the recordings took place. This transcription has yielded about 3000 lines of text. This text was hand tagged for part of speech, agreement morphology and class of nouns. While this corpus is small by the standards of corpus linguistics, it nonetheless provides sufficient information to estimate the distribution of features in highly frequent Tsez nouns.

4.2 Noun External Distributional Properties in the Corpus

As mentioned above, unique agreement for every class is only seen on vowel initial verbs and adjectives in Tsez. These verbs and adjectives make up only a small proportion of total verbs and adjectives in the dictionary (27% of verbs and 4% of adjectives). There are three possibilities concerning how this noun-external information is distributed in speech to children. First, it could be that this small proportion is reflected in the input, and hence that noun external cues to noun class are uncommon. Second, it could be that this proportion is even smaller in the input because the words exhibiting agreement are infrequent, making the use of noun external cues to noun class even more difficult. Finally, it could be that these vowel initial verbs and adjectives are highly frequent, thus providing robust noun external distributional cues to noun class.

To address this issue, we calculated the total number of verb and adjective tokens exhibiting agreement and compared it to the total number of verbs and adjectives. While the majority of verb types but only a minority of adjective types showed agreement (60% of verbs, 35% of adjectives), the majority of both verb and adjective tokens did show agreement (84% of verbs, 77% of adjectives).

Table 6: Proportions of verbs and adjectives that show overt agreement

	Agreeing Verbs	Agreeing Adjectives
Dictionary	27%	4%
Corpus Types	60%	35%
Corpus Tokens	84%	77%

These results, seen in Table 3, show that the agreeing forms are highly frequent, and thus that there are robust noun external distributional cues to noun class in the input to the learner of Tsez. Moreover, these cues are more frequent than would be expected given the distribution of vowel initial words in the overall Tsez lexicon.

4.3 Noun Internal Distributional Properties in the Corpus

Just as Plaster et al looked for noun internal regularities in the list of Tsez nouns from the dictionary, we wanted to look for such regularities in the nouns that children are exposed to. To do this, a list of nouns found in the corpus was compiled and tagged for morphophonological and semantic features similar to those used by Plaster et al. Decision trees were built using the unsupervised learning algorithm C4.5 in Weka, a machine learning toolkit (Witten & Frank 1998). Many similar features were found to be present in the child directed speech as in the dictionary, although there were some differences. Basically, three types of features were found to be useful in classifying nouns: biological semantic features (male, female, animate), other semantic features (paper, clothing) and morphophonological features (first/last segment). A summary of the most useful features for assigning words to each class, along with the predictive probabilities of each feature is found in Table 7:

Table 7: Predictive Features on Tsez Nouns in Child Directed Speech

Class	Biological Semantic	Other Semantic	Phonological
1	male human p(Cl1|male) =.99 p(male|Cl1) = .99	-----	-----
2	female human p(Cl2|female) = .99 p(female|Cl2) = .22	paper, clothing p(Cl2|cue) = .52 p(cue|Cl2) = .04	-----
3	animate p(cl3|animate) = .98 p(animate|cl3) = .13	-----	b- initial p(Cl3|b-) = .51 p(b-|Cl3) = .10
4	-----	-----	r-initial p(Cl4|r) = .61, p(r|Cl4) = .09 i final p(Cl4|-i) = .54, p(-i|Cl4) = .41

Now that we've established that, typewise, predictive features do exist for every class in the Tsez learner's input, it is important to show that these features appear frequently on nouns. It is important to note here that the phonological cues found to be predictive are identical to the agreement morphemes for these classes, but these are simply segments on the nouns *not* agreement morphemes, which are never present on nouns. The homophony is probably not accidental, and further work could address why this homophony between noun internal and noun external distributional information exists. An analysis of the corpus showed that out of 114 noun types heard, 24% had predictive features on them, and out of 1189 noun tokens heard, 39% had predictive features[8].

4.4 Correlation of Information Types

At this point we've shown that both noun external distributional properties and noun internal distributional properties are widely available to the Tsez learner. The EOH only requires that noun external distributional information be available for the classes to be acquired, but the HH requires not only that both noun external and noun internal distributional information are available, but that they are seen together. Therefore it is necessary to ask, how often does the Tsez acquiring child come across pairings of nouns with predictive features (noun internal distributional information) and agreement (noun external distributional information). Corpus analysis revealed that such cooccurence was quite frequent: 100% of class 1 nouns occurring with agreement also had predictive features[9], 52% of class 2 nouns, 51% of class 3 nouns and 45% of class 4 nouns.

Overall, the corpus analysis showed that both noun external and noun internal distributional properties are widely available to Tsez acquiring children, and are often available together. Thus the available input is consistent with that required by both the EOH and the HH. We must next address whether children's use of noun internal distribution mirrors adults' (that is, the

[8] These and other counts exclude proper names, which may decrease both the proportion of nouns with predictive features and the proportion of nouns with agreeing features seen with agreement, if the natural gender of the referent of a proper name can be though of as a predictive feature on the noun.

[9] This is perhaps trivial as all nouns in Class 1 denote male humans

distribution of this information in the input), supporting the EOH, or differs, supporting the HH. Additionally we will determine whether use of noun internal distributional information in general reflects a rule or probability based system.

5 Investigating Noun Class Acquisition in Tsez

The previous section established that the Tsez learner has available both noun external and noun internal distributional information for every noun class. As all the information necessary for either the EOH or the HH to hold is present, it is necessary to test the other predictions of these hypotheses: when children are able to use noun internal properties to classify nouns and whether they use them in proportion to their distribution in the input. In order to test sensitivity to the properties characteristic of groups of nouns in each class, classification of both frequent and novel nouns with combinations of the predictive features found above was elicited from adult and child speakers.[10]

5.1 Materials

The words used for classification were either real nouns that had the predictive featues or certain combinations of the features or nonce words invented to have these features. Table 8 shows the

---

[10] A pilot version of this task was conducted in summer of 2008 using features predicted by the decision tree in Plaster et al, and the task was revised both methodologically and in terms of the features on the words that were used in 2009. Only the results of the 2009 study will be reported here.

features that the different words had for each target class. A list of the words used can be found in Appendix A.

Table 8: Feature combinations on words used in classification task

Class	Biological Semantic	Other Semantic	Phonological	2 agreeing	2 conflicting
1	male human (3/3)	-----	-----	-----	male human & ɣ- initial (0/3) male human & b-initial (0/3)
2	female human (3/3)	paper (3/3) clothing (3/3)	ɣ- initial (3/3)	female human & ɣ-initial (0/3)	female human & r-initial (0/3)
3	animate (3/3)	-----	b- initial (3/3)	animate & b-initial (3/3)	animate & r-initial (2/3) animate & i final (0/3)
4	-----	-----	r-initial (3/3) i final (3/3)	r-initial & i final (2/3)	b-initial Cl4 real words (3/0)

Numbers in parentheses indicate the number of real and nonce items with these features (or feature combinations), separated by a /

Words either had a biological semantic feature, another semantic feature, a phonological feature, two features agreeing for class or two features predicting different classes. In the case of real words in Class 4 with conflicting features, they were actually in Class 4 but had the phonological cue (b- initial) for Class 3. The real words were frequent words either from the corpus of Tsez child directed speech or Tsez words whose translations were frequent in English child directed speech when the right combination of features wasn't available on Tsez words in the corpus. The nonce words were invented to conform to Tsez phonotactics and were checked with a native speaker to be sure they were not real words. Nonce words which had no predictive semantic or phonological information (other than the predictive value that comes from lacking certain features) were also included in order to be able to compare noun class assignment based on predictive information to that without.

The features selected had differing degrees of reliability as determined by the conditional probability of the feature given the class and by the conditional probability of the class given the feature. These differences will be important to keep in mind when considering whether the utility of noun internal distributional information is rule based or probability based, as well as when making specific predictions about classification when features make conflicting predictions. Table 9 summarizes the predictive information for each feature in the form of conditional probabilities for the class in question, given the feature and vice versa.

Table 9: Conditional probabilities of Class given feature used in experiment

Class	Biological Semantic	Other Semantic	Phonological
1	male human p(Cl1|male) =.99 p(male|Cl1) = .99	-----	-----
2	female human p(Cl2|female) = .99 p(female|Cl2) = .22	paper, clothing p(Cl2|cue) = .52 p(cue|Cl2) = .04	-----
3	animate p(cl3|animate) = .98 p(animate|cl3) = .13	-----	b- initial p(Cl3|b-) = .51 p(b-|Cl3) = .10
4	-----	-----	r-initial p(Cl4|r) = .61, p(r|Cl4) = .09 i final p(Cl4|-i) = .54, p(-i|Cl4) = .41

5.2 Predictions

Adults

When classifying real words, adults should make correct classifications regardless of the features on the nouns, as the classification for these words should be stored in their lexicons. When classifying nonce words, we expect adults to use the same cues that were predictive for words in the naturalistic speech examined in the corpus experiment. The distribution of classification when these cues are present will help to determine whether they are employed in a rule based or probability based system. Under a rule based system we would expect all words with a given feature to be classified according the the rule associated with that feature. Under a probability based system we would expect the distribution of nouns to classes to shift towards the class predicted by the feature, where the degree of skew is determined by the conditional probability of a given class given the feature in question. When classifying nonce words without cues, we will see whether the classification is determined by one default class or a default distribution mirroring the distribution of words without these cues into classes in the lexicon, further speaking to the question of whether classification based on noun internal information is rule based or probability based.

Children

The EOH predicts that children will perform similarly to adults with respect to the probabilistic nature of the cues available. This means that children should classify nonce words the same way adults do, and that if the cues on real words do affect their classification (perhaps in the case where a word is not well known), this should also follow the same principles that nonce word classification does. In particular, the EOH predicts that noun internal distributional properties are tracked later in development, at a point when the lexicon has full representations for both the form and meaning of each noun, thus the distribution of these properties in the intake should match the distribution in the input.

The HH in turn predicts that children's classification could differ from that of adults, as they would depend on noun internal distributional properties that are available from the very beginning of lexical acquisition. While some of these properties could be the same as those used by adults, it is possible that some would differ. For example, if children are able to track phonological information about words in conjunction with agreement morphology, these class internal regularities could be used even before the child knows the meanings of the words. A similar effect could be found if children find meaning an unreliable property to track early on in lexical acquisition. A learner can be fairly certain of the phonological form of a word that has been used, but may require more experience with that word to become as confident in the meaning. Thus in this case the distribution of noun internal information in the intake may differ from what is measurable in the input.

In summary, if adults and children pattern the same way in their use of noun internal cues, this would support the EOH, though perhaps not provide evidence to argue against the HH. However, if adults and children differ, exhibiting n difference between the input and the intake in children, we would have good reason to believe that despite the highly regular nature of the noun external information, both internal and external distributional properties are used to acquire a noun class system. Additionally, if use of noun internal distributional information by both adults and children appears to shift probabilities from a baseline distribution of nouns into classes, we would have good reason to believe that this information is used in a probability based system rather than a rule based system.

This work extends on the past work that found children favoring phonological over semantic information in the following ways (MacWhinney 1978; Karmiloff-Smith 1979; Perez-Pereira 1991; Mills 1985, 1986; Rodina 2009). First, in Tsez the biological semantic information has been shown to be more statistically reliable than the phonological information, unlike some of the cases in past work (i.e Mills 1985,1986). Thus it remains unclear what to expect when these two types of information conflict. Second, none of these studies directly compare adult and child performance on the classification of nonce words, with conflicting cues or otherwise. Finally, none of the past studies examined the behavior of adults and children on nonce forms with no predictive information. This is important in determining if a certain cue has an effect on classification or if speakers are simply relying on default class probabilities, and also in determining whether there is a rule based system or a probability based system employed, both in the classification of words with predictive noun internal information and those without.

5.3 Task

The task exploited the fact that vowel initial verbs show agreement. Verbal agreement in Tsez is absolutive agreement, thus intransitive verbs agree with the agent and transitive verbs agree with the patient. The verb *eat* is vowel initial in both the intransitive *-iš* and the transitive *-ac'o* and so will show agreement. During the task a native Tsez speaking assistant manipulated a flat paper figure on a page of a book. The page had various objects drawn it, arranged pseudo randomly such that no page had all items from one class and no page was without something potentially edible. The child was trained on the task and told to tell the figure first to start eating (using intransitive -*iš*) as this would show agreement with the eater. Then the figure would move around the page and the assistant would point out and name each object. The child would tell the character to eat it or not using the transitive *-ac'o,* and in doing so show agreement with the thing being eaten. Thus the child thought the task was about determining was what edible. In telling the character what it should or shouldn't eat, participants were expected to use agreement and to implicitly classify the nouns in question when doing so. A sample page is shown in Figure 1, and an idealized transcript of a trial is found in Table 10.

Figure 1: Sample Experimental Items

kid (*girl)* *Class 2* *Semantic Cue*	buq (*sun)* *Class 3* *Phonological Cue*	k'uraj (*onion)* *Class 4* *no Cue*	zamil (*nonce)* *Class 3* *Semantic Cue*

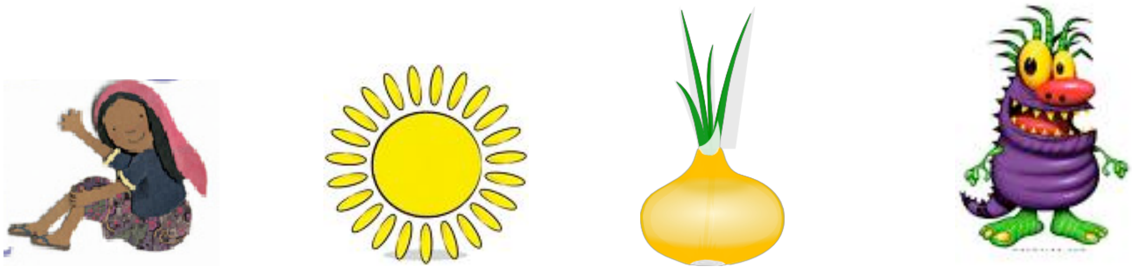


Table 10: Model Trial

Speaker	Linguistic Stimuli/Response	Action
Assistant	kid girl(Class 2)	*explains task, points to human character and labels it*
Child	sis, q'ano, ɬono, **j**-iš one two three CL2-eat *One two three, Eat!*	*instructs character*
Assistant	buq sun(Class 3)	*points to sun, labels it*
Child	buq **b-**ac'xosi aanu sun CL3-eat-pres.part neg pro *isn't eating the sun*	*instructs character/describes scene*
Assistant	k'uraj onion(Class 4)	*points to onion, labels it*
Child	k'uraj **r**-ac'o onion CL4-eat *eat the onion*	*instructs character/describes scene*

Speaker	Linguistic Stimuli/Response	Action
Assistant	zamil nonce(target Class 3)	*points to nonce animal and labels it*
Child	zamil **b**-ac'xosi aanu zamil CL3-eat-pres.part neg pro *isn't eating the Zamil*	*instructs character/describes scene*

5.4 Participants

Participants were native Tsez speakers living in Shamkhal and Kizilyurt, Dagestan[11]. They were recruited with the help of a local Tsez speaking assistant who knew Tsez speaking families in the area. Data from 10 young children (ages 4-7), 12 older children (ages 8-12) and 10 adults was included in the analysis below. Because the number of children available to participate was rather small, we created large age ranges to test, creating a basic distinction between older and younger children. Subjects were tested either alone in a room with the experimenter and a native Tsez speaking assistant, and sometimes were accompanied by parents, relatives or other friends who were instructed to keep silent during the experiment, with some encouraging remarks being allowed when the child being tested was especially shy.

20 additional children and 3 additional adults participated but were excluded from the final analysis for one of 3 reasons: (1) because other people were present during the experiment and

[11] The Tsez speakers in these communities are immersed in a bi- or tri-lingual environment (with Russian and Avar), as these are settlements outside of the traditional Tsez speaking region. Access to the Tsuntinsky region, where Tsez is the native language, is highly restricted by the Russian government, meaning that at the time of this work the region was inaccessible. However, Tsez, not Russian or Avar, is still the main language spoken in the homes of the subjects in question, and was the language child subjects spoke to one another when observed outside of the experimental context.

prompted the subject with answers (2 children, 1 adult), (2) because they failed to use agreeing forms on a majority of the items (4 children), or (3) because they failed to classify 8 out of 10 very frequent words correctly (14 children, 2 adults). (3) was used as an exclusion criterion because a common strategy for participants was to classify all of the words in one class (either Class 3 or Class 4). The latter two categories of behavior are puzzling, as they do not seem to show the classification or agreement system that the speaker has. This is apparent in that participants exhibiting this behavior were observed using proper agreement when conversing outside of the task. Because of the extension of this behavior to real, known words in the task, it it clear that it is not just a reflex of some ‘default’ class. Rather, it appears that this is some kind of task induced strategy used by certain participants, and while it doesn’t show much about the classification of individual items, it might highlight a part of the classification system that has not yet been discussed. One possibility is that these participants were classifying everything as if the noun were *picture* (which is in Class 3), or some other noun that would serve the same function but is in Class 4. This would mean that instead of classifying each item, they were just using a form that agreed with *picture* or some Class 4 noun. Alternatively, some mechanism may be employed under special circumstances to override actual class assignment and show apparent agreement with nothing in particular. This is no doubt an interesting puzzle but falls outside of the scope of the current work.

5.5 Results

Classification data from the experiment was analyzed as follows. For each item type (i.e. nonce word with semantic feature 'female' or real word with phonological feature 'b- initial'), the proportion of items put in each class was calculated for each age group. For example, for young children, for the item type 'nonce words with semantic feature 'female'', 4 % were put in Class 1, 52 % in Class 2, 22 % in Class 3 and 22 % in Class 4. This yielded a unique distribution of proportions of nouns assigned to each class for each item type and each age group. The differences between these distributions were quantified using Jensen-Shannon Divergence (J-S divergence), a metric for quantifying the difference between sample distributions (Lin 1991). By comparing the differences between distributions for each cue type, we could determine which cues caused the distributions to change, and to what degree. What follows is a summary of the main findings from comparing these distributions. A full presentation of every item type and age group, as well as an explanation of the calculation of the J-S divergence used to quantify the differences between them can be found in Appendices B & C. The data was analyzed in this way instead of through using t-tests or ANOVAs to compare the proportion of nouns in a given class given a set of cues because those tests were deemed inappropriate to compare the shift of classification across a set of classes. That is, it mattered not only that a cue could raise or lower the proportion of nouns in a given class, but also how the distribution was skewed with the introduction of a given cue.

In analyzing the results, classification of real words was compared to the words' actual class. Classification of nonce words with cues was compared to a base distribution of classification of nonce words without cues. When talking about the classification of real words, we'll refer to what proportion of words of each item type were assigned to the words actual class (the class of the word agreed upon by native speaker consultants). When talking about the classification of nonce words we'll refer to what proportion of the words were assigned to the target class (the class that the cue on the item most strongly predicts) as compared with the proportion of words assigned to that class when no cue was present. For example, the target class of a nonce word referring to a female human would be Class 2, and so we look at nonce words with female referents to see if more are assigned to Class 2 when the cue is present, than nonce words without this cue.

5.5.1 Classification of real words

We expect that if speakers know the class of a given word and the task is effective in eliciting this classification, the classification data found in the experiment will match the class agreed upon by native speaker informants. That is, speakers should assign the actual class to each word. For most word types, this is what we found (Table 11).

Table 11: Percentage of real words of each type correctly assigned to actual class

	Biological Semantic	Other Semantic	Phonological	No Cue	Conflicting
Young Children	79	71	84	77	**42***
Older Children	86	58	94	78	**47***
Adults	87	75	92	86	71

The * indicates that the J-S divergence between the classification distributions of words with cues conflicting with actual class assignment and the classification distribution of words in this class without the conflicting cue was in the top 10% of the distribution of all J-S divergences for real words.

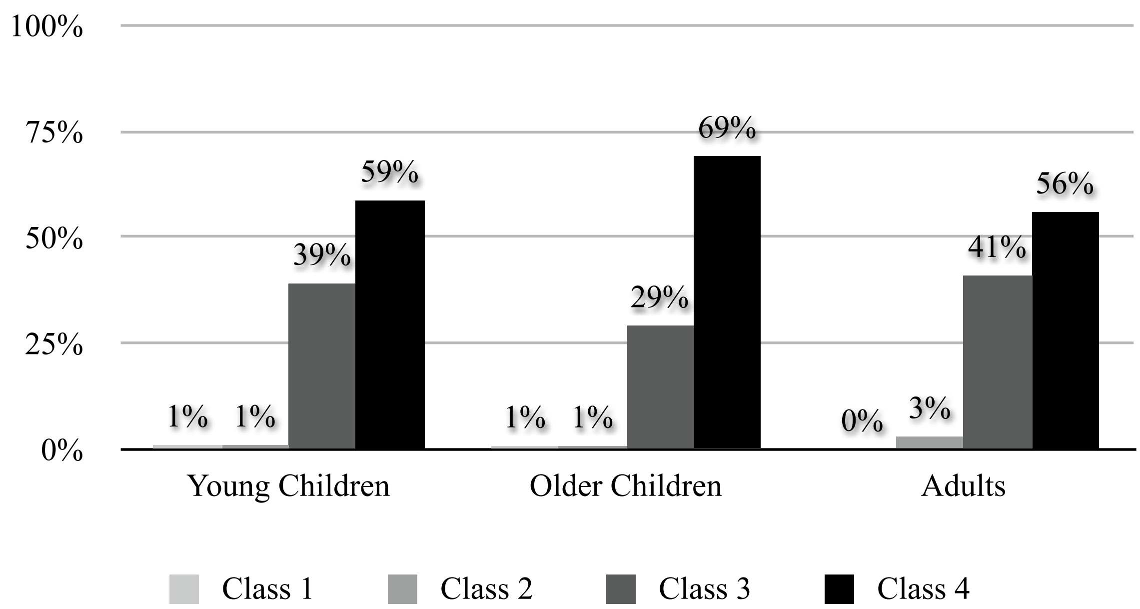
However, there are several things to point out in this data. First of all, in no case was classification perfect. This most likely reflects noise from this being experimental task, rather than an imperfection in the classification of speakers as a group.

This caveat aside, we can see that all age groups performed very well on classifying words with semantic and phonological or no apparent cues to their class. However, when cues conflicted with the actual class of the words, it appears that children in both age groups were influenced by this conflicting information. In all cases, the conflicting information was a phonological cue to another class, while the word was a member of a different class. For example, *recenoj* (ant) is in class 3, but begins with [r], which is a cue for class 4. This means that for children, the phonological cue to a given class tended to outweigh the linguistic experience that the child would have with the word.

5.5.2 Classification of nonce words without cues

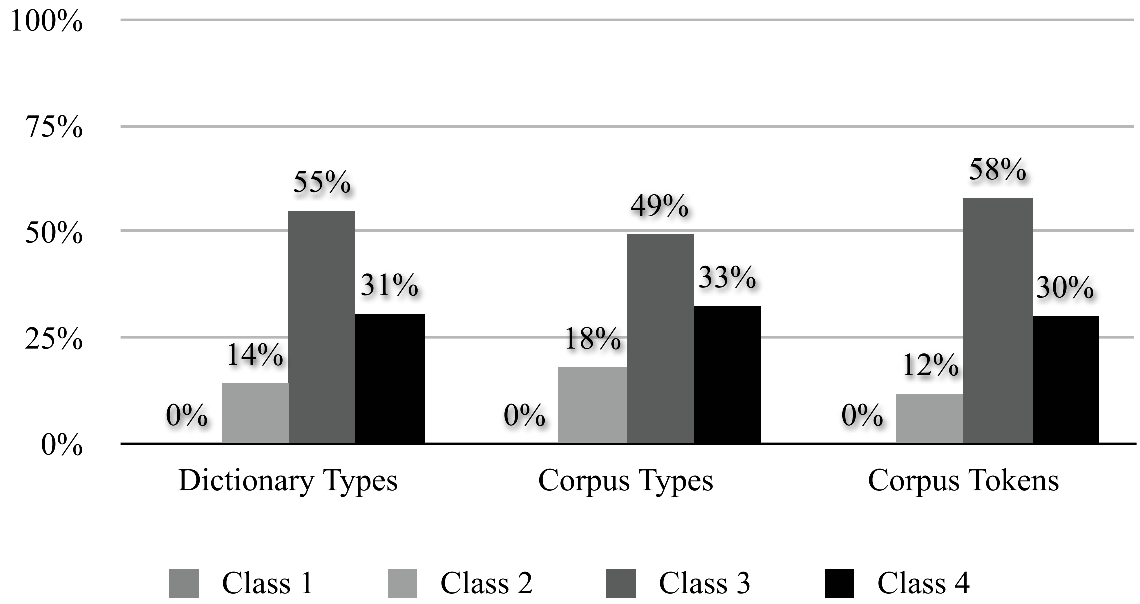
Next we will consider the classification of nonce words with no predictive features. It must be noted, however, the the lack of predictive feature is in itself a predictive feature (e.g. not being a male human means the noun is not in Class 1) There are two ways that nouns without predictive features could be treated: they could be assigned to one default class or they could be distributed across classes based on the relative probabilities that any noun would be in any class. The results of this classification task are seen in Figure 2.

Figure 2: Classification of Nonce Words Without Cues: Percentage of words assigned to each class by age group



Across all age groups, nouns appear to be distributed according to a probability distribution of noun classes. Exactly what determines the shape of this distribution is unclear: is it based on type or token frequencies or something more complex? In Figure 3 we can look at the type frequencies of noun class in the dictionary and to that of type and token frequencies of noun classes in the corpus.

Figure 3: Frequencies of nouns without predictive cues in the dictionary and corpus



While the default classification distribution doesn't precisely map onto any of these distributions, it is important to keep in mind that the unnatural nature of the task could be adding complexity to the distribution that might not be there in the most naturalistic setting, as well as the fact that lack of a predictive feature is also a predictive feature. Other factors could also be shaping this distribution, and forthcoming modeling work addresses this issue (Gagliardi, Feldman & Lidz 2012). Whatever factors determine the precise nature of this distribution, it is clear that classification in the absence of noun internal and noun external information reflects some baseline probability of nouns into classes, probably modulated by the absence of certain predictive features, not a default assignment rule. It is this baseline distribution that is important to keep in mind when examining the effect that predictive cues have on the classification of nonce words. As we will see, these cues only work to skew this distribution in the direction indicated by the predictiveness of the cue, not as rules assigning nouns to classes.

5.5.3 Classification of nonce words with cues

Unlike with the classification of real words, where we expected the majority of words to be assigned to their actual class, when looking at the classification of nonce words we expect words to be classified according the the distribution outlined above, unless the cues on the words have an effect on the classification. That is, if the cues on the nonce words influence their classification we expect to see a modulation from the default distribution. In Table 12 we can see the proportion of words assigned to the target class (the class the cue is predicted to signal).

Table 12: Percentage of Nonce words of each type correctly assigned to target class

	Biological Semantic	Other Semantic	Phonological	Conflicting
Young Children	54	**8***	61	**38***
Older Children	65	**9***	63	53
Adults	53	**23***	61	55

The * indicates that the J-S divergence between the classification distributions of words with these cues types he classification distribution of words with other cues to these classes was in the top 10% of the distribution of all J-S divergences for real words.

This data must be interpreted not only as the proportion of words assigned to the target class, but also in terms of how much this proportion varied from the default classification. We can see that semantic and phonological cues are effective in getting the majority of words assigned to the target class by all age groups. For Classes 1 and 2, this is is also very different from the default distribution. While the difference is not as extreme for Classes 3 and 4, where the majority of the words ended up by default, examination of the data by class shows that the vast majority of words end up there when the relevant cues are present, many more than when no cues are present. Full profiles of the classification for each cue type by class can be seen in Appendix B.

It is more difficult to see how other semantic information is used. Remember that other semantic cues were only tested for Class 2. Children do not appear to use this information at all, as the 8% and 9% of nonce words assigned to Class 2 with the information do not significantly differ from the 1 % of cueless words assigned to Class 2 (The J-S divergence between these distributions does not fall in the top 10% of all J-S divergences). For adults on the other hand, while the 23% of words with the other semantic cue assigned to Class 2 is not the majority, it does differ significantly from the proportion of words assigned to this class without this cue.

Finally, the effect of conflicting information is also apparent. Nonce words with conflicting information were those that had cues to two different classes - semantic and phonological. In all cases, the semantic information was a statistically better predictor of class, as the probability that a real word with that cue will be in the class is higher than the probability that a word will be in the class predicted by the phonological cue (probabilities that a word will be in a given class are in Table 13, copied from Section 5.1 above).

Table 13: Conditional probabilities of class given cue used in experiment

Class	Biological Semantic	Other Semantic	Phonological
1	male human p(Cl1|male) =.99 p(male|Cl1) = .99	-----	-----
2	female human p(Cl2|female) = .99 p(female|Cl2) = .22	paper, clothing p(Cl2|cue) = .52 p(cue|Cl2) = .04	-----
3	animate p(cl3|animate) = .98 p(animate|cl3) = .13	-----	b- initial p(Cl3|b-) = .51 p(b-|Cl3) = .10
4	-----	-----	r-initial p(Cl4|r) = .61, p(r|Cl4) = .09 i final p(Cl4|-i) = .54, p(-i|Cl4) = .41

Thus the class of the the semantic cue can be thought of as the target class for these examples. Despite the higher predictive power of the semantic cues, young children failed to use them to assign nouns to the target classes, and relied more heavily on the less predictive phonological information. The conflicting phonological information did not appear to have this effect on the older children and adults.

5.6 Discussion of Results

Overall, we found that adults and children will classify nouns in this task. This classification is influenced by properties on the nouns themselves. Semantic and phonological cues are used by both adults and children to classify nonce words in a manner consistent with the predictions these types of cues make. When these cues make conflicting predictions, or when it conflicts with the actual class of a real word, young children are more likely to use phonological information, despite the fact that this information is statistically less predictive. Finally, the classification of nonce words with and without predictive cues follows some distribution, influenced by both the noun internal distributional cues (or lack thereof), as well as a baseline distribution of nouns into classes.

6 General Discussion

The EOH predicted that children would have access to statistical regularities of inherent noun properties late in the acquisition of noun classes, but that when they did their generalizations should mirror the adult ones. The HH predicted that children would be able to access statistical regularities from the onset of lexical acquisition, but that their initial use of these regularities could differ from adults, as the first available regularities might be different from those used by adults. While these results do not test children young enough to speak to the question of whether statistical regularities are used by children from the very beginning of lexical acquisition they do appear to point towards the HH for the following reason.

First, while both children and adults classify novel nouns based on noun internal properties, the features they take advantage of do not have the same statistical reliability in the input. That is, when all of these features are fed into a decision tree building algorithm, the biological semantic ones can classify with 100% accuracy whereas the phonological ones do not do as well. Yet, children appear to weigh the phonological cues more heavily when determining the class of a novel noun. This highlights a distinction in the input and the intake. Some characteristic of the intake mechanism puts a higher value on phonological rather than semantic information. There are three reasons this could be so, all pointing towards the utility of noun internal distributional information in very early acquisition. First, phonological properties of words are available to a child who might be able to track phonological features and their relation to agreement morphemes long before knowing the meaning of the words in question. Second, once a child is actually learning words, the phonological form is reliably as it sounds, whereas the meaning of the word in question may not be as easy to grasp the first few times the word is heard. Third, the learner could have a bias to track phonological information rather than semantic information stemming from either the early observation that phonological information is more useful, or from a bias to prefer phonological information over semantic information.

All three of these possibilities raise interesting questions about the nature of the developing lexicon, in particular what information can be stored and accessed as part of a lexicon before words have well defined (or any) semantics attached to them. This is an important question, and not one that can go unanswered in precisely characterizing the process of noun class acquisition. However, for the purposes of the current paper it suffices to say children rely on the kind of information that is available at the earliest stages of lexical development, and that they do so despite this information being less statistically reliable in the environment. That is, the information they use, the intake, does not match the information that is available in the input. Recall that the HH predicted that this was possible, which the EOH predicted that the intake should match the input and the behavior by adults. Thus this observation supports the HH over the EOH.

6.1 Specifying the role of the predictive noun internal distributional information

Future computational modeling efforts will allow us to look precisely at the effect of predictive cues on the classification of nonce nouns. As alluded to above, it appears as if speakers are classifying based not only on the predictive cues that a noun has, but based on the joint probability of classification given these cues, some prior or baseline probability for a noun to be in each class, and perhaps other factors as well. By modeling exactly what these probabilities are we will get predictions for how each word type would be classified by such a system, and compare these (and thus our model) to actual classification, gaining a better understanding of what kind of categorization system this predictive information is playing a role in. Additionally, we will be able to investigate the question of what information is available to the early learner and how we would predict classification based on this information, shedding light on the nature of the filter on the input and the early stages of the acquisition of noun classes.

6.2 Mechanisms and further thoughts

Although this work points toward the hypothesis that children pay attention early on to both noun external and noun internal distributional properties, it hasn't addressed the precise mechanism that would require these two types of information in conjunction. There are two ways we are currently investigating exactly what the properties of this mechanism might be.

Studies using miniature artificial languages (Braine 1987, Frigo & MacDonald 1998, Gerken et al 2002, Gerken et al 2005) have shown that in order for learners to discover multiple lexical classes and generalize to new items, a subset of the items in each class must have some regularity among them. That is, in order for learners to discover classes in these artificial languages, the item external distributional information alone is not sufficient to induce classification, and some item internal distributional information must also be available to the learner. While these studies were done using very small toy languages, the striking similarities between the information necessary for adult and infant subjects to acquire classes in the laboratory and the information available to and used by children acquiring noun classes in natural language are very suggestive. Current work focuses on expanding these artificial language results to make the toy languages more like natural ones in an effort to see if the pattern still holds. In this way, we may begin to understand precisely what kinds of information are used, and what kind of mechanism could make use of them.

Computational models of noun class acquisition will also be important in investigating this mechanism further. By building explicit models of the acquisition process we will be able to see what kinds of mechanisms take advantage of both kinds of information, and under what conditions these models perform better than models that use only noun external information. Modeling will also allow us to test predictions about why children use phonological information more than semantic: because it is available earlier or because it is more reliably detected. Finally, building explicit models about the processes at work in language acquisition will give us further, testable hypotheses about how noun class acquisition proceeds.

6.3 An Extension to Verb Classes

As mentioned earlier, current models of English irregular verb classes are insufficient to capture noun class behavior. These models are based on the premise that there are as many verb classes as there are clusters of verbs behaving in one way or another, and within these clusters one can extract phonological and/or semantic regularities among verbs that characterize the majority of the group. In the case of noun classes, large groups of nouns cluster together with respect to how they behave (noun external distributional information), but the clusters of nouns with semantic or phonological similarities only make up a small subsection of each class. Pinker's Words and Rules model (1991), which posits that English speakers have a rule for regular past tense and a number of memorized exceptions, doesn't appear appropriate for this kind of data. While it might be possible to posit a few 'regular rules' based on predictive semantic information and perhaps a default rule, the majority of the lexicon would have to be listed as exceptions to these rules. Moreover, children do not appear to be using semantic features as if they were 'regular rules' or a 'default rule', and rather appear to be classifying nouns probabilistically. Yang's Rules and Competition model(2002) posits that there are many rules that compete to form the past tense of any given verb. While this might cover the words that can be classified based on noun internal distributional information, it would depend on rules that classify only one word to cover at least a third of the lexicon, and rules that classify only two words for another third (compare with Plaster et al's decision tree rules). Hay and Baayen (2005) propose a probabilistic system in which verbs are classified based on how similar they are to other verbs. This seems partially alignable to noun class systems, in that novel nouns are classified based on shared properties with other nouns. However, the architecture of this system misses the overarching class structure: nouns with a given feature don't simply act like other nouns with this feature, they act like a whole class of nouns that may or may not have that feature. It is unclear both how this generalization would be captured in such a model, especially when the majority of a class has no apparent features in common. While none of these models appear as a good fit for our data on noun classification, it is possible that our hypotheses regarding noun classification might be capable of capturing irregular verb classes and this topic deserves future investigation.

6.4 Concluding Remarks

In this paper we have looked at the acquisition of noun classes, a problem that allows us to differentiate between the *input*, or the information available to a learner in the environment, and the *intake*, the information that a learner actually makes use of. While we have not directly investigated the early acquisition of noun classes, we have investigated the predictions that two hypotheses about this acquisition make regarding later behavior in noun classification. In doing so we have been able to draw inferences regarding what information children make use of when discovering noun classes and determining which nouns are in which class. In the acquisition of Tsez noun classes we find that input and intake do differ. While Tsez acquiring children appear to make use of both noun external and noun internal distributional information, their use of noun internal distributional information is selective. Instead of using semantic cues, which both adults and statistical models find to be the most reliable information, children use less reliable phonological information. This finding suggests that the earliest stages of noun class acquisition depend not only on noun external properties such as agreement, which define the classes, but also regularities among nouns in a class. It also allows us to understand more about what kind of mechanism lies behind noun class acquisition, and to set up further studies to probe the exact character of this mechanism. Additionally, this investigation allowed us to examine whether noun classification in the absence of unambiguous external distributional information follows assignment rules or some underlying distribution of nouns into classes, and our results supported the latter hypothesis.

7. References

Ambridge, B., Pine, J. M., Rowland, C. F. & Jones, R. L. (2009). Incremental learning of verb- and construction-semantics as a solution to the 'no negative evidence' problem. *Cognitive Science*, 33, 1301-1316.

Braine, M. D. S. (1987). What is learned in acquiring word classes – a step toward an acquisition theory. In B. MacWhinney (ed.), *Mechanisms of language acquisition.* Hillsdale, NJ : Erlbaum.

Chomsky, N. (1959). Review of Verbal Behavior by B.F. Skinner, *Language,* 35, 26-58.

Chomsky, N. (1965) *Aspects of a Theory of Syntax*, Cambridge, MA: MIT Press.

Comrie, B. and Polinsky, M. (1999). Some Observations on Class Categorization In Tsez. In H. Van den Berg (ed). *Studies in Caucasian Linguistics: Selected Papers of the Eighth Caucasian Colloquium* .The Netherlands: Universitat Leiden.

Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences*, 4, 597-650.

Cyr, M and R. Shi. (2009). Grammatical Gender Categorization in 17-month-old Infants. Conference Presentation. BUCLD34.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge: MIT Press.

Frigo, L. & J. McDonald. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language,* 39, 218–45.

Gagliardi, A., Feldman, N. & Lidz, J. (2012) Psychocomputational approaches to the acquisition of noun classes in Tsez. Talk to be given at the Pyschocomputational models of Language Acquisition conference, Portland, OR.

Gerken, L. A., Wilson, R., Gómez, R. L. & Nurmsoo, E. (2002). Linguistic category induction without reference : the importance of correlated cues. Unpublished manuscript, University of Arizona.

Gerken, L., R. Wilson, and Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, 32, 249-268.

Gold, E. M. (1967). Language identification in the limit. *Information and Control,* 10, 447-474.

Halle, M. & Mohanan, K.P. (1985). Segmental Phonology of Modern English. *Linguistic Inquiry* 16:1, 57-116.

Harris, Z. (1951). *Methods in Structural Linguistics*. Chicago: The University of Chicago Press.

Hay, J. B. & Baayen, R. H. (2005). Shifting paradigms: gradient structure in morphology, *Trends in Cognitive Sciences* 9, 342-348.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

Karmiloff-Smith, A. 1979. *A functional approach to child language: A study of determiners and reference*. Cambridge: Cambridge University Press.

Khalilov, M. Sh. (1999). *Tsezsko-russkij slovar'*. Moscow: Academia.

Lewis, J., & Elman, J. (2001). A connectionist investigation of linguistic universals: Learning the unlearnable. In R. Amos, C. Bradford, C. Jefferson, and D. Meyers (Eds.), *Proceedings of the Fifth International Conference on Cognitive and Neural Systems.* Center for Adaptive Systems and the Department of Cognitive and Neural Systems, Boston University.

Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Trans. on Information Theory*, 37(1), 145--151.

Lidz, J. & Pearl, L. 2007. "Language Learning Through Selection." *Science* 315:332.

MacWhinney, B. (1978). The Acquisition of Morphophonology. *Monographs of the Society for Research in Child Development*, Vol 43, No 1/2, Blackwell.

Mills, A.E. (1985). The acquisition of German. In D. Slobin (Ed) *The crosslinguistic study of language acquisition, volume 1*, 141–254. Hillsdale, NJ: Lawrence Erlbaum.

Mills, A. E. (1986.) *The Acquisition of Gender: A Study of English and German*. Berlin: Springer.

Pérez-Pereira, M. (1991). T*he acquisition of gender: What Spanish children tell us. Journal of Child Language*. 18(3), 571-590

Pinker, S. (1979). Formal models of language learning. *Cognition* 7, 217–83.

Pinker, S. (1984). *Language Learnability and Language Development* . Cambridge, Mass.: Harvard University Press.

Pinker, S. (1991). Rules of Language. *Science*, 253, 530–535.

Plaster, K., M. Polinsky and B. Harizanov. (2009). Noun Classes Grow on Trees: Noun Classification in the North-East Caucasus. To appear.

Polinsky, M., and D. Jackson. (1999). Noun classes: Language change and learning. In B. Fox, D. Jurafsky, and L. A. Michaelis (eds). *Cognition and Function in Language*, 29-58. Chicago: The University of Chicago Press.

Quinlan, J. R. (1993). C4.5: Programs for machine learning. San Mateo, CA: Morgan Kaufmann Publishers.

Rodina, Y. (2009). Semantics and Morphology. The Acquisition of Grammatical Gender in Russian. Unpublished Doctoral Dissertation, University of Tromsø.

Rumelhart, D. E., & McClelland (1986). On learning the past tenses of English verbs. In McClelland, J. L., Rumelhart, D. E., and the PDP research group (Eds.) *Parallel distributed processing: Explorations in the microstructure of cognition. Volume II*. 216-271. Cambridge, MA: MIT Press.

Tucker, G.R., W.E. Lambert and Rigault, A. (1977). *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. The Hague: Mouton

Witten, I. H. and Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. 2nd Edition, San Francisco, CA: Morgan Kaufmann.

Yang, C. D. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Appendix A: Nouns Used in Classification Experiment

Word Type	English	Tsez
Nonce, 1, Conflicting Cue	novel man	ɣasi
Nonce, 1, Conflicting Cue	novel man	ɣeža
Nonce, 1, Conflicting Cue	novel man	banu
Nonce, 1, Conflicting Cue	novel man	ɣušon
Nonce, 1, Conflicting Cue	novel man	bino
Nonce, 1, Conflicting Cue	novel man	buma
Nonce, 1, Semantic Cue	novel man	cina
Nonce, 1, Semantic Cue	novel man	kirop
Nonce, 1, Semantic Cue	novel man	melu
Nonce, 2, Agreeing Cues	novel woman	ɣeħu
Nonce, 2, Agreeing Cues	novel woman	ɣunik
Nonce, 2, Agreeing Cues	novel woman	ɣina
Nonce, 2, Conflicting Cue	novel woman	riɫu
Nonce, 2, Conflicting Cue	novel woman	rak’o
Nonce, 2, Conflicting Cue	novel woman	ruja
Nonce, 2, Phonological Cue	novel food	ɣobar
Nonce, 2, Phonological Cue	novel object	ɣuto
Nonce, 2, Phonological Cue	novel food	ɣaɬa
Nonce, 2, Universal Semantic Cue	novel woman	kuna
Nonce, 2, Universal Semantic Cue	novel woman	haba
Nonce, 2, Universal Semantic Cue	novel woman	sohaq
Nonce, 2, Idiosyncratic Semantic Cue	novel paper	molo
Nonce, 2, Idiosyncratic Semantic Cue	novel clothing	lemin
Nonce, 2, Idiosyncratic Semantic Cue	novel paper	mačum

Nonce, 2, Idiosyncratic Semantic Cue	novel clothing	kenu
Nonce, 2, Idiosyncratic Semantic Cue	novel paper	ħidar
Nonce, 2, Idiosyncratic Semantic Cue	novel clothing	zubu
Nonce, 3, Agreeing Cues	novel animal	bazu
Nonce, 3, Agreeing Cues	novel animal	budu
Nonce, 3, Agreeing Cues	novel animal	biʕan
Nonce, 3, Conflicting Cues	novel animal	yugi
Nonce, 3, Conflicting Cues	novel animal	resu
Nonce, 3, Conflicting Cues	novel animal	riga
Nonce, 3, Conflicting Cues	novel animal	čoħi
Nonce, 3, Conflicting Cues	novel animal	rola
Nonce, 3, Conflicting Cues	novel animal	t’awi
Nonce, 3, Phonological Cue	novel food	beɬo
Nonce, 3, Phonological Cue	novel food	baka
Nonce, 3, Phonological Cue	novel food	bidan
Nonce, 3, Semantic	novel animal	zamil
Nonce, 3, Semantic	novel animal	seno
Nonce, 3, Semantic	novel animal	kiru
Nonce, 4, Agreeing Cues	novel food	rubi
Nonce, 4, Agreeing Cues	novel object	reħi
Nonce, 4, Agreeing Cues	novel food	rabi
Nonce, 4, Phon. Cue -i	novel food	tali
Nonce, 4, Phon. Cue -i	novel object	joni
Nonce, 4, Phon. Cue -i	novel object	q’omi
Nonce, 4, Phon. Cue r-	novel object	rega
Nonce, 4, Phon. Cue r-	novel food	ruɬo
Nonce, 4, Phon. Cue r-	novel food	rinaɣ
Nonce, No Cue	novel food	miraj
Nonce, No Cue	novel food	lesi

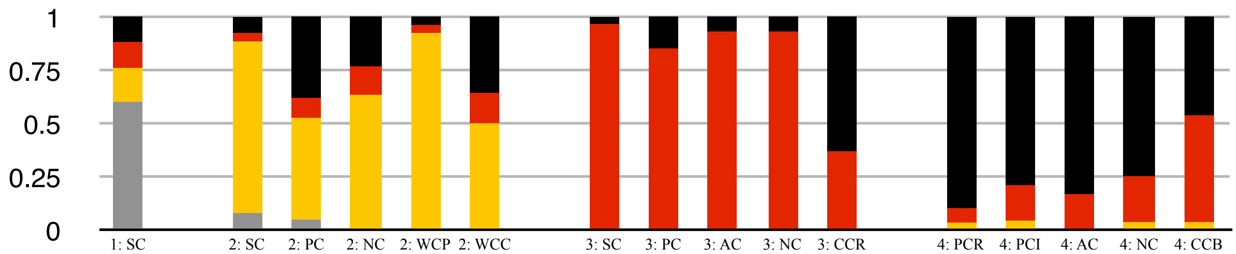
Nonce, No Cue	novel food	kola
Nonce, No Cue	novel food	nola
Nonce, No Cue	novel food	kela
Nonce, No Cue	novel food	šiwa
Nonce, No Cue	novel food	dero
Nonce, No Cue	novel object	norib
Nonce, No Cue	novel food	žewu
Nonce, No Cue	novel food	nawe
Real, 1, Semantic Cue	baby	k’ak’a
Real, 1, Semantic Cue	boy	uži
Real, 1, Semantic Cue	father	baba
Real, 2, No Cue	salt	cijo
Real, 2, No Cue	door	ac
Real, 2, No Cue	cheese	izu
Real, 2, Phonological Cue	stone	ɣuɬ
Real, 2, Phonological Cue	milk	ɣaj
Real, 2, Phonological Cue	pants	ɣeł’o
Real, 2, Universal Semantic Cue	woman	ɣana
Real, 2, Universal Semantic Cue	girl	kid
Real, 2, Universal Semantic Cue	mother	eni
Real, 2, Idiosyncratic Semantic Cue	letter	kaɣat
Real, 2, Idiosyncratic Semantic Cue	shirt/dress	ged
Real, 2, Idiosyncratic Semantic Cue	underwear	turusik
Real, 2, Idiosyncratic Semantic Cue	hat	šapka
Real, 2, Idiosyncratic Semantic Cue	book	t’ek
Real, 2, Idiosyncratic Semantic Cue	newspaper	gazit
Real, 3, agreeing cues	fish	besuro
Real, 3, agreeing cues	snake	bikori

Real, 3, agreeing cues	sheep	be’ƛ’ɣu
Real, 3, conflicting cues	sea	raɬad
Real, 3, conflicting cues	ant	recenoj
Real, 3, no cue	apple	heneš
Real, 3, no cue	potato	hek’u
Real, 3, no cue	bread	magalu
Real, 3, phonological cue	sun	buq
Real, 3, phonological cue	cherry	ba’li
Real, 3, phonological cue	finger	baša
Real, 3, semantic cue	chicken	onoču
Real, 3, semantic cue	cow	zija
Real, 3, semantic cue	cat	k’et’u
Real, 4, conflicting cue	outhouse	butka
Real, 4, conflicting cue	flag	bairaq
Real, 4, conflicting cue	ring	basčiqow
Real, 4, no cue	onion	k’uraj
Real, 4, no cue	soup	čorpa
Real, 4, no cue	eye	ozura
Real, 4, Phon. Cue -i	water	ɬi
Real, 4, Phon. Cue -i	porridge	qiqi
Real, 4, Phon. Cue -i	window	aki
Real, 4, Phon. Cue r-	hand	reƛ’a
Real, 4, Phon. Cue r-	butter	riɬ
Real, 4, Phon. Cue r-	key	reka
Real, 4, Agreeing Cues	trash	rešoni
Real, 4, Agreeing Cues	cradle	rikini

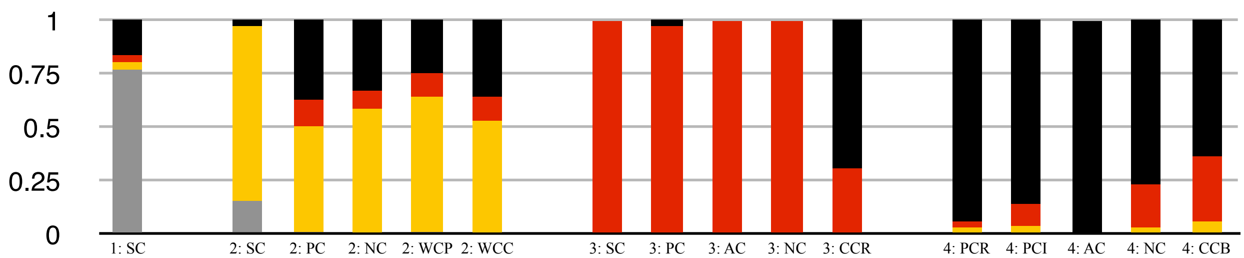
Appendix B: Full Classification Results for each Item Type

Figure B1: Classification of Real Words

a. Younger Children



b. Older Children



c. Adults

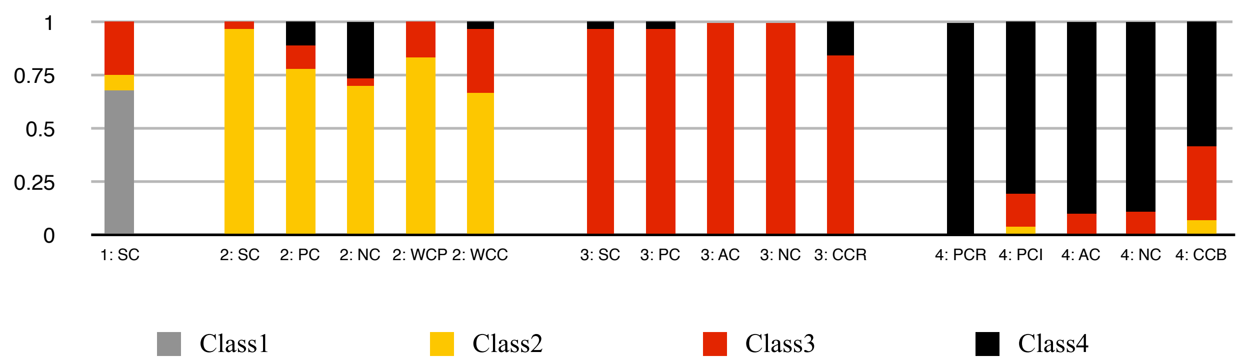
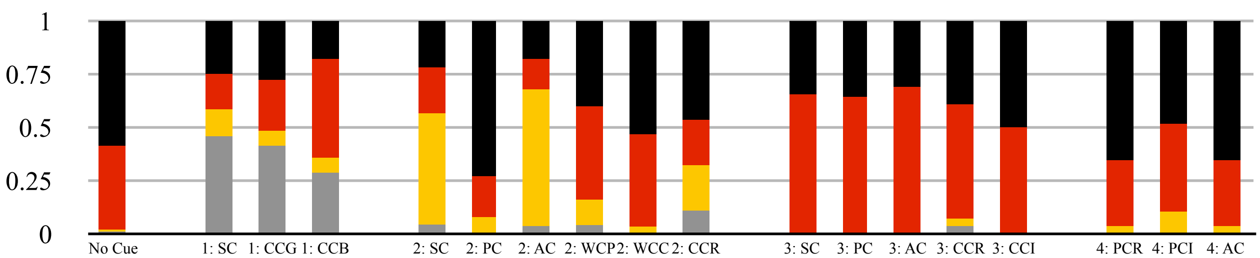


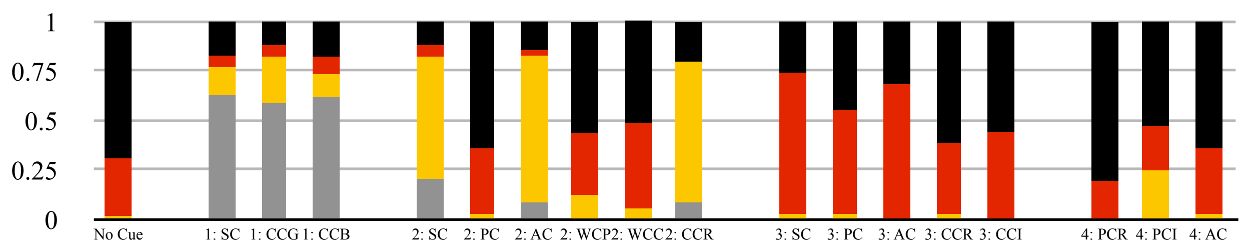
Figure B1: Each bar in the figure corresponds to a set of test items, grouped above by target class. The colors in the bars correspond to the proportion of words from this set assigned to the target class. Speakers generally assign nouns to the class they belong in, though when predictive information for two classes is in conflict, children tend to use phonological information and adults semantic. The item type that each bar corresponds to can be found in Table B1.

Figure B2: Classification of Nonce Words

a. Young Children



b. Older Children



c. Adults

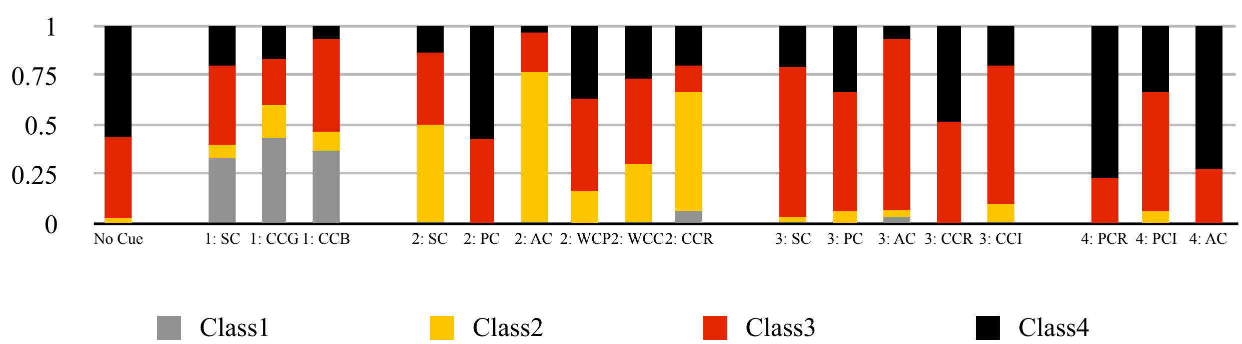


Figure B2: While nonce words show more noise, there is an evident effect of biological semantic cues on all groups, though only adults appear to use other semantic cues. Phonological cues are used, except those for class 2 (probably related to an misrepresentation of the frequency of this cue in the input). When semantic and phonological information conflict children appear most likely to use phonological information and adults semantic (not when this information is the non working phonological cue for class 2). Codes for each item type can be found in Table B1.

Table B1: Codes for Item Type

Code	Cue Type	Cues (class associated with cue)
1: SC	Biological Semantic Cue	male (Cl1)
2: SC	Biological Semantic Cue	female (Cl2)
3: SC	Biological Semantic Cue	animate (Cl3)
2: WCP	Other Semantic Cue	paper (Cl2)
2: WCC	Other Semantic Cue	clothing (Cl2)
2: PC	Phonological Cue	b- initial (Cl3)
3: PC	Phonological Cue	ɣ- initial (Cl2)
4: PCR	Phonological Cue	r- initial (Cl4)
4: PCI	Phonological Cue	-i final (Cl4)
2: AC	Biological Semantic and Phonological Cues	female & ɣ- initial (Cl2)
3: AC	Biological Semantic and Phonological Cues	animate & b- initial (Cl4)
4: AC	Biological Semantic and Phonological Cues	r-initial & -i final (Cl4)
1: CCG	Conflicting Cue	Class 1 semantic cue with Class 2 Phonological Cue
1: CCB	Conflicting Cue	Class 1 semantic cue with Class 3 Phonological Cue
2: CCR	Conflicting Cue	Class 2 semantic cue with Class 4 Phonological Cue
3: CCR	Conflicting Cue	Class 3 word (real) or Class 3 semantic cue with Class 4 Phonological Cue
3: CCI	Conflicting Cue	Class 3 word (real) or Class 3 semantic cue with Class 4 Phonological Cue
4: CCB	Conflicting Cue	Class 4 word (real) with Class 3 Phonological Cue
NC	No Cue	No Predictive Cue

Appendix C: Jensen-Shannon Divergence

The results discussed above were analyzed as follows. For every set of words with a given feature or set of features, the proportion of words assigned to each class was calculated. This meant that for each set of words we had a distribution of noun class assignment for each age group. In order to determine whether distributions were really different from one another, the Jensen-Shannon (JS) divergence was calculated between each relevant pairing of distributions (i.e. all the sets with target class 2). JS divergence is a symmetrized form of Kullback-Leibler divergence, which is a measure of how much one distribution differs from another (Lin, 1991). The equation for calculating JS Divergence is shown in Equation 1.

Equation C1:

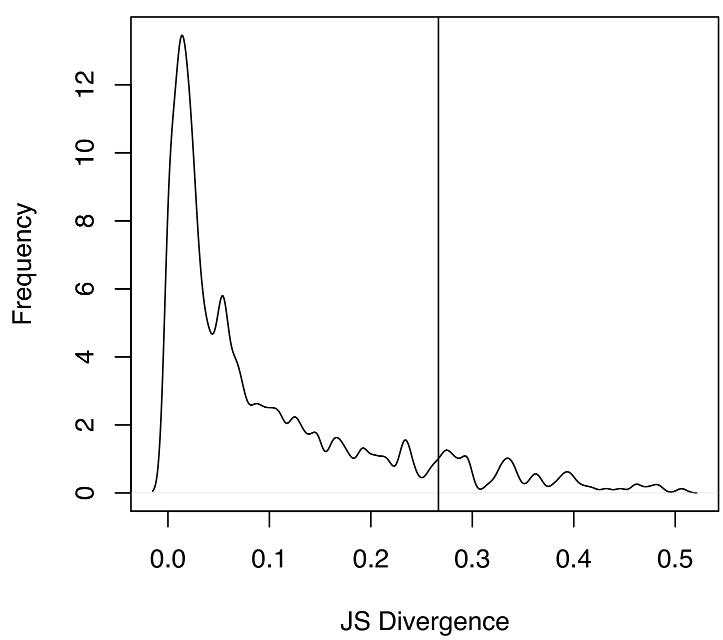
$$D_{JS}(P||Q) = \frac{1}{2} D_{KL}(P||M) + \frac{1}{2} D_{KL}(Q||M)$$

$$\textit{where } M = \frac{1}{2}(P+Q)$$

$$\textit{and } \quad D_{KL}(P||M) = \sum P(i)\log(P(i)/M(i))$$

This resulted in a distribution of possible JS divergences for the data under consideration (Figure C1).

Figure C1: Distribution of JS Divergences



The JS divergence between a pair of sets of interest (i.e. adults' use of a phonological cue for Class 3 vs young children's use of the same cue) was examined with respect to the resulting distribution of JS divergences to determine where it fell in the distribution. The divergences between distributions considered 'different' below were those that fell in the top 10% of the distribution.

The comparisons across groups in the paper do not directly reference the JS divergences for a given cue, class and group. Instead, they compare the proportion of nouns assigned to the actual class (real words) or target class (nonce words) for a given cue type by each group. These proportions are compiled from all of the distributions for a given group and cue type (i.e. young children's use of phonological cues for classes 2, 3 and 4) and then compared to one another. The JS divergences between the distributions that these proportions are compiled from (e.g. all the distributions based on young children's use of conflicting cues vs. all of those based on adults' use of conflicting cues) tell us whether these compiled proportions reflect real differences. The following patterns emerged from this analysis:

(1) Classification of nonce words with phonological or semantic cues for classes 1, 2 and 3 reliably differed from classification on nonce words with no cues, but this classification did not differ across groups
(2) Classification of of nonce words with conflicting cues differed from classifications of words with only phonological or semantic cues for both child groups but not the adult group
(3) Classification of real words with conflicting cues differed from classification of real words for only the group of younger children
(4) Classification of nonce words with other semantic cues did not differ from classification of words with no cues for either child group, but did for the adult group

Thus, the differences in the proportions presented in the data in the main body reflect actual differences in the classification of nouns by speakers in the experiment.