Statistical insensitivity in the acquisition of Tsez noun classes

Annie Gagliardi, Jeffrey Lidz

Language, Volume 90, Number 1, March 2014, pp. 58-89 (Article)

Published by Linguistic Society of America
DOI: 10.1353/lan.2014.0013

STATISTICAL INSENSITIVITY IN THE ACQUISITION OF TSEZ NOUN CLASSES

Annie Gagliardi
Harvard University

Jeffrey Lidz
University of Maryland, College Park

This article examines the acquisition of noun classes in Tsez, looking in particular at the role of noun-internal distributional cues to class. We present a new corpus of child-directed Tsez speech, analyzing it to determine the proportion of nouns that children hear with this predictive information and how often this is heard in conjunction with overt information about noun class agreement. Additionally, we present an elicited production experiment that uncovers asymmetries in the classification of nouns with versus without predictive features and by children versus adults. We show that children use noun-internal distributional information as a cue to noun class out of proportion with its reliability. Children are biased to use phonological over semantic information, despite a statistical asymmetry in the other direction. We end with a discussion of where such a bias could come from.*

Keywords: language acquisition, noun classes, Tsez, input, intake

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 task of language 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 analyzing 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, inferable from linguistic experience (in the absence of explicit constraints on the character of linguistic structure) and attempts to bring this into alignment with how children develop (e.g. Lewis & Elman 2001, Ambridge et al. 2009).

* This research was supported by the NSF Graduate Research Fellowship Program and by NSF IGERT 0801465. We are grateful to Magomed I. Magomedov, Arsen Abdulaev, Patimat Abdulaeva, and Xayrat Gamzatova for assistance in Dagestan, and to Maria Polinsky, Colin Phillips, the UMD Project on Children's Language Learning, and the UMD Cognitive Neuroscience of Language Laboratory for assistance and helpful discussion.

In the current article, we consider a case that is at odds with both perspectives: the acquisition of noun classes by children in Tsez, a Nakh-Dagestanian language spoken by about 6,000 speakers in the Northeast Caucasus.[1] On the one hand, children seem to fare considerably worse in this task than linguists do, missing obvious generalizations; on the other hand, they draw different conclusions from the statistical information than a purely distributional learner would. Whereas the linguist armed with some simple tools of distributional analysis can master the noun class system of a language like Tsez in a relatively brief time, children apparently struggle with such systems 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 both linguists and computational models is informative 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 that is utilized by the learning mechanism in building a grammar. This distinction gives us some insight into the particular distributional analyses that children are prepared to engage in, as well as those that they may be predisposed to avoid. While our particular focus here is on Tsez, work on which is a contribution in itself since it is an understudied and endangered language, addressing these issues also makes an important, independent contribution to the study of language acquisition in general.

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 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 children make use of this information as a cue to noun class, we 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 (i) whether the language in question had noun classes, (ii) how many classes there were, and (iii) which class each noun used with agreement belonged to. 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 were not, the language in question would not have a noun class system.

With these two types of information (highly regular and probabilistic) available in principle to the learner, we can ask what information the learner makes use of when going through the 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 that feeds forward into the construction of a grammatical system? While it may look like there

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

is ample evidence for the existence and structure of the noun classes in the input, the portion of this evidence that the learner uses 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 by the child (Fodor 1998, Valian 1999, Pearl & Lidz 2009). 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 of the information 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 article, we look at learners with a developing system of noun classes. By looking at how this developing system differs from the adult system we can glean information about (i) how the learner thinks nouns are organized into classes and (ii) what portion of the available information the learner must have used to arrive at this state. These two pieces of evidence allow us to draw inferences about the discovery of noun classes earlier in development. First, we examine what information is available in the input by constructing and analyzing a corpus of child-directed Tsez speech. Focusing on noun-internal information, we go on to look at what adult and child speakers are sensitive to when classifying novel nouns. Despite a statistical asymmetry in the input where semantic information is a more reliable predictor of class than phonological information, the children, but not adults, appear to be biased toward phonological over semantic information. This suggests differences between the available input and the intake used by children to acquire the noun class system.

We first detail what noun-internal and noun-external information looks like in Tsez, the Nakh-Dagestanian language we use to investigate the acquisition of noun classes (§2), and then lay out several hypotheses relating noun-external and noun-internal distributional information to the acquisition of noun classes, as well as give an overview of related work (§3). A new corpus of child-directed Tsez is presented in §4, together with 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. The key observation of the article is found in §5: 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. We then show how the experimental findings support the view that both noun-internal and noun-external information is critical to noun class acquisition, relating them back to the input/intake distinction (§6). Finally, we put forward several hypotheses accounting for the existence of this distinction in the acquisition of noun classes.

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[2]) 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

[2] 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 correct, since 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 article we use the term noun class, since it suggests no primacy of certain correlating features over others.

classifier systems the influence of a noun's class is limited to the noun phrase.[3] Noun classes can be fully characterized by the noun-external distributional properties such as the agreement paradigm or syntactic behavior that defines the class, and partially characterized according to noun-internal distributional properties, the nonrandom distribution of characteristics of the nouns that make up each class. As mentioned above, these two types of information could be used in noun class acquisition.[4]

2.1. NOUN-EXTERNAL DISTRIBUTIONAL PROPERTIES. Noun classes are defined as groups of nouns that pattern the same way with respect to noun-external properties such as agreement. Languages differ as to where this agreement is seen (Corbett 1991). Some languages are limited to DP-internal agreement, 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 many as twenty (Fula) (Corbett 1991).

Our particular focus here is on Tsez, which 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 verbs, adjectives, and adverbs[5] (Table 1).

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'

TABLE 1. Tsez singular noun class agreement.

Thus the agreement prefix that appears on adjectives modifying class 1 nouns is the null prefix, for class 2 it is [j], class 3 [b], and class 4 [r]. The same set of prefixes is used on verbs, adjectives, and adverbs. Plural agreement prefixes and some forms of both personal and demonstrative pronouns also vary by noun class,[6] but there is considerable syncretism in these paradigms, making them less reliable markers of class (Tables 2–4).

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 2. Tsez plural noun class agreement.

[3] This article focuses on noun class systems, but similar arguments could be applied to the acquisition of classifier systems (see e.g. Hu 1993).

[4] Certain types of verb classes might be superficially characterized in a similar way—members of a class share both 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.

[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.

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

		class 1 (singular)	class 2–4 (singular)	class 1 (plural)	class 2–4 (plural)
1st person	absolutive	di		eli	ela
	oblique	dā-		elu-	ela-
	genitive	dej		eli, eliz	
2nd person	absolutive	mi		meži	meža
	oblique	debe-, dow-		mežu-	meža-
	genitive	debi		meži, mežiz	

Table 3. Tsez personal 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

Table 4. Tsez demonstrative pronouns.

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. Because only singular agreement provides reliable evidence for the existence of four classes, we restrict our attention to singular agreement marking for the remainder of the article.

2.2. Noun-internal distributional properties. If we suspect that noun-internal distributional information is important for the acquisition of noun classes, it is imperative to determine whether languages have, for each class, some feature or set of features characteristic of the nouns in that class. The results of many typological surveys are resoundingly positive: every noun class system appears to have some regularity in the way at least a subset of nouns are classified (Corbett 1991). For the acquisition researcher investigating whether 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 the child makes use of this information during acquisition. Below we look at the noun-internal distributional information that characterizes Tsez noun classes.

A summary of the classes based on traditional descriptions of the language (Comrie & Polinsky 1999) is found in Table 5.

class 1	class 2	class 3	class 4
(13% of nouns)	(13% of nouns)	(41% of nouns)	(34% of nouns)
all male humans	all female humans	all other animates	many other things
only male humans	many other things	many other things	

Table 5. Summary of Tsez noun classes; percentages reflect the percentage of the nouns in each 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 to any other class. Not reflected in the 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, since all female humans belong in class 2. Unlike class 1, however, the majority of class 2 is made up of inanimate or abstract nouns. Class 3 is the largest class, and, while it contains all animate, nonhuman 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 and colleagues (2013) investigated the kinds of information that characterize nouns in Tsez and found that rules referring to additional properties of the nouns themselves (noun-internal information) could classify many more nouns in the dictionary. They took the set of nouns from a Tsez dictionary (Khalilov 1999) and tagged them for possibly predictive features. These included semantic features, such as animacy and various physical and functional properties; phonological features, such as first and last segments and morphemes, and number of syllables; and formal features, such as the declension class. The result was a feature vector for each noun that included values for every possible feature. 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 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 and colleagues were able to accurately classify about 70% of Tsez nouns in the dictionary. Semantic features, both those referencing properties like animacy and humanness and those referencing physical properties like being stone or being a container, were found to be more predictive than formal properties like 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 and colleagues 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 in 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 formulate testable hypotheses about what information makes up the intake, and how this information may be used. There are two senses in which these information sources could be used: by adults both to represent their noun class systems and to classify novel nouns, and by children to acquire the system of classes and to classify novel nouns as they learn them.

In the discussion that follows, we 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 and do use noun-external distributional properties when processing sentences, and presumably this information is also diagnostic of the class of novel nouns for adult users. 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 since 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% likely 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 [ɪŋ[+velar]] form. However, neither the existence of the *i–a* ablaut nor the [ɪŋ[+velar]] form is predictive of the other (e.g. *spit–spat*, *think–*thank*). Different analyses model the relationship between these irregular properties and class membership in different ways: as a class of exceptions to a regular rule (Pinker 1991), as multiple rules acting over small classes of words that tend to have phonological similarities (Halle & Mohanan 1985, Yang 2002), or as 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 expand on this observation in §6 and examine whether our conclusions about noun classification may bear on irregular verb classes.

Returning to the classification of novel nouns by adult Tsez speakers, we do not know at this point whether noun-internal 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. Taken in their simplest form, these two alternatives appear to make distinct predictions for the classification of nonce words. A rule-based 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 probabilistic system predicts that if nouns with a certain feature have some probability distribution across classes, and if this feature is observed on a novel word, the probability that the novel word is in a given class will be proportional to a probability computed from the following: (i) the probability distribution of nouns with this cue, (ii) the prior probability of each class, and (iii) the probabilities associated with any other predictive features this noun contains. By specifying what this probability is we can precisely model the classification of novel words. This modeling falls outside the scope of the current article and is addressed in other work (Gagliardi et al. 2012). What is important for this article is that classification based on noun-internal predictive information will work either deterministically, as in a rule-based system, or probabilistically, as in a probabilistic system. Of course, we should also mention that it is possible that speakers would use a distribution of probabilities to determine a rule-based system (see e.g. Yang 2004, Pearl 2011). If this were the case, we might expect to see adults using noun-internal information deterministically and children using it probabilistically, or not at all.

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 must by definition have some default classification rule for such nouns, whereas a probabilistic system could 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.

At the end of this section discussing whether predictive noun-internal information is used deterministically or probabilistically, it seems important to point out that while we invoke a 'rules vs. probabilities' dichotomy in our characterization of this problem, this division is orthogonal to the familiar debate from the 1980s about whether linguistic computation is symbolic (e.g. Rumelhart & McClelland 1986, Pinker & Prince 1988, and the vast literature those papers spawned). Our use of probabilities here is entirely within a symbolic architecture, as, for example, in Labov 1969, Sankoff 1971, Booth & Thompson 1973, Jelinek 1990, Yang 2004, and Pearl & Weinberg 2007, inter alia, and therefore does not bear on the debates about connectionist versus symbolic architectures, as they were played out in the domain of past-tense morphology and its acquisition.

3.2. Acquisition of noun classes. No matter the precise way in which noun-internal distributional information works, in order 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. In order to acquire noun classes the learner must (i) notice that the language has noun classes, (ii) determine how many classes there are, and (iii) determine which nouns go in which classes. Below we outline two hypotheses about 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 about earlier steps is likely to have led to the behavior we observe.

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. First, the child could simply use noun-external distributional information in the beginning to discover classes and classify nouns as they are 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 morphemes competed for the same 'verb agreement' slot in the paradigm. Only noun-external information is nec-

[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.

essary for such a system to work—the existence of noun-internal information would not hinder this process, but would not be necessary for classes to be acquired, either. Since this system does not rely on noun-internal distributional information, in order 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 has 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.

The 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 cooccurrence of these subclasses of nouns with class-dependent noun-external distributional information. At a certain stage, the learner 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 he observed 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 noun-internal information and later uses the noun-external information. Within this hypothesis there are, of course, several others. These subhypotheses reflect the fact that there are several types of noun-internal information that characterize Tsez noun classes. That is, we can ask whether children use both semantic and phonological information, or only one or the other. Moreover, we can ask whether certain types of semantic information, such as linguistically common classification criteria like animacy or natural gender, are preferred as compared with less common criteria like being 'made of paper' or 'used for clothing'.

The two hypotheses concerning the role of noun-internal information in the acquisition of noun classes make different predictions about the differences between input and intake in noun class acquisition. If children use only noun-external information, we predict that they may be insensitive to 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. Our reasoning is as follows: because 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, phonological and semantic features may not be equally well represented early in development, due to the fact that phonological features will be accessible to the learner earlier than semantic ones. While this would cause the intake to differ from the input, these discrepancies would not be expected to carry over to the more mature noun class system, which has been formed on the basis of cooccurrence of noun-external morphology alone. When these properties are eventually incorporated, they will be drawn from a mature lexicon, and will thus closely match the noun-internal distributional properties attended to by adults.

The predictions differ, however, if children acquire noun classes by tracking dependencies both on internal features among nouns and external agreement information. If this is the case, we might predict that they should be sensitive to noun-internal distributional properties from the earliest point at which they can track such dependencies. Since the lexicon is still being formed at this early stage, it is possible that the statistical regularities extracted early on will reflect not 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. That is, features that children can track earlier in development, phonological and morphological, will at least initially be of greater use to children than features that are encoded later on, such as semantics. This means that, as suggested above, we might see different sensitivity to different kinds of noun-internal information in children's noun classification (e.g. all predictive internal features, only crosslinguistically common semantic features, or only phonological features).

Given this characterization of the problem, we can specify four distinct possibilities concerning children's use of noun-internal features during noun class acquisition, summarized in Table 6. Note that these hypotheses are orthogonal to the question of whether noun-internal features are used in a rule-based or probabilistic fashion. That question probes what kind of classification mechanism makes use of noun-internal features, and this one probes whether those features are used in acquisition. Neither of these questions addresses debates surrounding the symbolic versus subsymbolic computational architectures.

hypothesis	features used in acquisition	prediction: When internal features are used in generalization …
external only	no noun-internal features	classification is in proportion to their statistical reliability.
all available internal	all available noun-internal features	classification relies on noun-internal features in proportion to their statistical reliability in the input throughout development.
common internal	only crosslinguistically common internal features	classification relies more on common semantic features than both (i) crosslinguistically uncommon features and (ii) phonological features.
only phonological	only phonological features	classification relies more on phonological than semantic features.

Table 6. Hypotheses and predictions for the use of noun-internal information in acquisition.

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 view that both are crucial for noun class acquisition. Work in French (Karmiloff-Smith 1979), Spanish (Pérez-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 suggest that children acquire noun classes by tracking morphophonological dependencies both among nouns and between nouns and agreement from a very early age. Unfortunately, this work does not directly address the questions posed by the hypothesis outlined above, since 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 directly investigate (i) the statistical distribution of noun-internal and noun-external distributional information in the input as it can be measured in a corpus of child-directed speech, (ii) how well adult and child speakers' sensitivity to this information aligns with its statistical reliability, and (iii) 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 several 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. Were we examining this issue in English or another commonly studied language, we might have these corpora available to us, which we would use to rigorously examine how much of this information is available in the input that learners receive. Since, however, we are looking at this problem in Tsez, we had no such corpus. To address this issue, as well as provide data for future work, we created a corpus of child-directed speech in Tsez. Once we have characterized the information that the learner is exposed to, we can investigate hypotheses about how this information is used.

4.1. The corpus. Over a period of one month, ten hours of child-directed speech were recorded during normal daily interactions between a mother, aunt, and older sister of two twenty-month-old Tsez-acquiring children in Shamkhal, Dagestan. Roughly six 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 3,000 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 seen only 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 singular 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).

	AGREEING VERBS	AGREEING ADJECTIVES
DICTIONARY	27%	4%
CORPUS TYPES	60%	35%
CORPUS TOKENS	84%	77%

TABLE 7. Proportions of verbs and adjectives that show overt singular agreement.

These results, seen in Table 7, 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. These numbers are similar when the percentage of verbs and adjectives showing agreement is broken down by class, showing that agreement is commonly seen for each class.

4.3. NOUN-INTERNAL DISTRIBUTIONAL PROPERTIES IN THE CORPUS. Just as Plaster and colleagues (2013) 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 and colleagues. Decision trees were built using the unsupervised learning algorithm C4.5 in Weka, a machine learning toolkit (Witten & Frank 2005). 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 (made of paper, used for 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 8.

CLASS	BIOLOGICAL SEMANTIC	OTHER SEMANTIC	PHONOLOGICAL
1	male human *p*(CL1|male) = 0.99 *p*(male|CL1) = 0.99	—	—
2	female human *p*(CL2|female) = 0.99 *p*(female|CL2) = 0.22	paper, clothing *p*(CL2|cue) = 0.52 *p*(cue|CL2) = 0.04	—
3	animate *p*(CL3|animate) = 0.98 *p*(animate|CL3) = 0.13	—	b-initial *p*(CL3|b-) = 0.51 *p*(b-|CL3) = 0.10
4	—	—	r-initial *p*(CL4|r-) = 0.61 *p*(r-|CL4) = 0.09 i-final *p*(CL4|-i) = 0.54 *p*(-i|CL4) = 0.41

TABLE 8. Predictive features on Tsez nouns in child-directed speech.

Now that we have 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.[8] An analysis of the corpus showed that out of 114 noun types heard, 24%

[8] 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 from a historical perspective, and further work could address why this homophony between noun-internal and noun-external distributional information exists.

had predictive features on them, and out of 1,189 noun tokens heard, 39% had predictive features.[9]

4.4. Correlation of information types. At this point we have shown that both noun-external distributional properties and noun-internal distributional properties are widely available to the Tsez learner. Thus this data is consistent with all of our hypotheses outlined above. There is sufficient noun-external information available that a child might be able to acquire the noun class system from this data alone. Abundant predictive noun-internal information is also available that could be used to augment this process. To use these two types of information together, however, they must cooccur. It is therefore necessary to ask how often the Tsez-acquiring child comes across pairings of singular nouns with predictive features (noun-internal distributional information) and singular agreement (noun-external distributional information). Corpus analysis revealed that such cooccurrence was quite frequent: 100% of class 1 nouns occurring with agreement also had predictive features,[10] as did 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 all hypotheses we have laid out as possibilities for noun class acquisition. We must next address whether children's use of noun-internal distributional information mirrors adults' (that is, the distribution of this information in the input), supporting the hypothesis that they rely on external information only to acquire noun classes, or differs, supporting the hypothesis that (i) both noun-internal and noun-external information is used, and (ii) that children's sensitivity to noun-internal features differs from their predictiveness in the input. If speakers use noun-internal distributional information when classifying novel nouns, our experimental data will allow us to determine whether use of this 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. Since all of the information necessary for any hypothesis to prevail 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, which kinds of properties they use, 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.[11]

5.1. Materials. The words used for classification were either real nouns that had the predictive features or certain combinations of the features, or nonce words invented to have these features. Table 9 shows the features that the different words had for each target class. A complete list of the words used can be found in Appendix A.

[9] These and other counts exclude proper names. This exclusion 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 thought of as a predictive feature on the noun.

[10] This is trivial since all nouns in class 1 denote male humans

[11] A pilot version of this task was conducted in summer of 2008 using features predicted by the decision tree in Plaster et al. 2013, and in 2009 the task was revised both methodologically and in terms of the features on the words that were used. Only the results of the 2009 study are reported here.

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)

TABLE 9. Feature combinations on words used in classification task. Numbers in parentheses indicate the number of real and nonce items with these features (or feature combinations), separated by a slash.

Words had either a biological semantic feature, another semantic feature, a phonological feature, two features agreeing for class, or two features predicting different classes. We included these three types of features since they allow us to examine whether speakers use semantic or phonological features, as well as whether different kinds of semantic features can be used (natural classes like animacy and artificial ones like 'made of paper'). 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 either frequent words 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 was not 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 ensure that they were not real words. Nonce words were said to have predictive semantic information when they were presented as referring to novel animals or human characters. Nonce words that 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 with 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 features are those summarized above in Table 8). In particular, biological semantic features were the most predictive of class, and phonological features were less so. These differences will be important to keep in mind when considering whether the use 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.

5.2. PREDICTIONS.

ADULTS. When classifying real words, adults should make correct classifications regardless of the features on the nouns, since 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 probabilistic system. Under a rule-based system, we would expect all words with a given feature to be classified according to the rule associated with that feature. Under a probabilistic system, we would expect the distribution of nouns to classes to shift toward 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 by 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 probabilistic.

Children. To outline the predictions we have for children in this classification task, we return to the summary of our hypotheses and their predictions (Table 10).

HYPOTHESIS	FEATURES USED IN ACQUISITION	PREDICTION: When internal features are used in generalization …
EXTERNAL ONLY	no noun-internal features	classification is in proportion to their statistical reliability.
ALL AVAILABLE INTERNAL	all available noun-internal features	classification relies on noun-internal features in proportion to their statistical reliability in the input throughout development.
COMMON INTERNAL	only crosslinguistically common internal features	classification relies more on common semantic features than both (i) crosslinguistically uncommon features and (ii) phonological features.
ONLY PHONOLOGICAL	only phonological features	classification relies more on phonological than semantic features.

Table 10. Hypotheses and predictions for the use of noun-internal information in acquisition.

If children rely solely on noun-external distributional information to acquire noun classes, we predict that they 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. If adults use all cues, so then should children. This is because if children use only noun-external distributional information to acquire noun classes, then 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, and thus the distribution of these properties in the child's lexicon should match the distribution in the input. Note that this prediction depends on children having acquired a noun class system AND having available the semantic and phonological features of nouns in this system. Since these children are four years old and older, the assumption that they are at this stage does not seem unreasonable.

If instead children use the combination of noun-internal and noun-external distributional information, we predict that children's classification could differ from that of adults. If they use all noun-internal distributional properties as they become available in development, some of the features could be the same as those used by adults, but it is possible that some would differ. As we outlined above, 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. The reverse is also possible. Since we can find crosslinguistic generalizations in some of the semantic information characterizing noun classes (such as natural gender), it is possible that children have an expectation that this information will be relevant and it is then the information they first pick up on, meaning that at least early on they will not be sensitive to crosslinguistically uncommon semantic features, or to phonological features. Finally, it is possible that children have a bias (either learned or inherent) to rely only on phonological information in development. All three of these hypotheses predict that 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 idea that noun-external information alone is used for noun class acquisition. However, if adults and children differ, with children exhibiting a difference between the input and the intake, we would have good reason to believe that despite the highly regular nature of the noun-external information, both noun-internal and noun-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 probabilistic rather than a rule-based system.

This work extends on past work that found children favoring phonological over semantic information (MacWhinney 1978, Karmiloff-Smith 1979, Mills 1985, 1986, Pérez-Pereira 1991, Rodina 2009) in the following ways. 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 without predictive information. In trying to determine whether a certain cue has an effect on classification, it is important to know how speakers classify nouns when no predictive information is available, since this allows us to see whether the classification pattern when that cue is present looks identical to classification when no cues are present. This is also important in determining whether a rule-based or probabilistic system is employed in the classification of words both with predictive noun-internal information and 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 theme. 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 on it, arranged pseudo-randomly such that no page had all of its items from just 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*), since this would show agreement with the agent (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 theme (the thing being eaten). Thus the child thought the task was about determining was what edible. In telling the character what it should or should not eat, participants were expected to use agreement and to implicitly classify the nouns in question when doing so. The experimenter recorded what agreement morpheme (and thus what class) was used in conjunction with each experimental item. The procedure was repeated for multiple pages, until all 107 items had been tested. A sample page is shown in Figure 1, and an idealized transcript of a trial is found in Table 11.

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

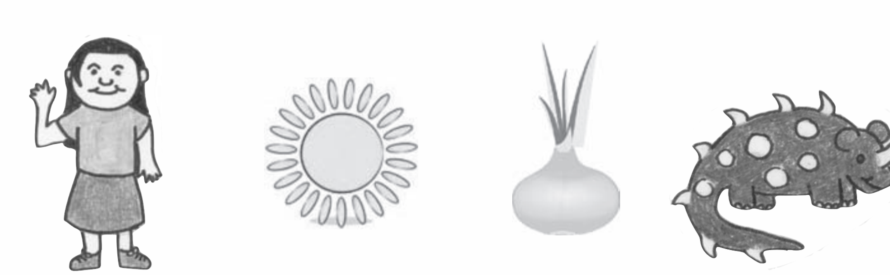

FIGURE 1. Sample experimental items.

SPEAKER	LINGUISTIC STIMULI/RESPONSE	ACTION
Assistant:	kid girl (class 2) 'girl'	*explains task, points to human character, and labels it*
Child:	sis, q'ano, łono, **j**-iš one two three II-eat 'One, two, three, Eat!'	*instructs character*
Assistant:	buq sun (class 3) 'sun'	*points to sun, labels it*
Child:	buq **b-**ac'-xosi aanu sun III-eat-PRS.PTCP NEG '*pro* isn't eating the sun'	*instructs character/describes scene*
Assistant:	k'uraj onion (class 4) 'onion'	*points to onion, labels it*
Child:	k'uraj **r**-ac'o onion IV-eat 'eat the onion'	*instructs character/describes scene*
Assistant:	zamil nonce (target class 3) 'Zamil'	*points to nonce animal and labels it*
Child:	zamil **b**-ac'-xosi aanu zamil III-eat-PRS.PTCP NEG '*pro* isn't eating the Zamil'	*instructs character/describes scene*

TABLE 11. Model trial.

5.4. PARTICIPANTS. Participants were native Tsez speakers living in Shamkhal and Kizilyurt, Dagestan.[12] They were recruited with the help of a local Tsez-speaking assistant who knew Tsez-speaking families in the area. Data from ten young children (ages four to seven), twelve older children (ages eight to twelve), and ten 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

[12] The Tsez speakers in these communities are immersed in a bi- or trilingual environment (with Russian and Avar), since 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 participants in question, and was the language in which child participants spoke to one another when observed informally outside of the experimental context.

younger and older children. Subjects were tested 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.

Twenty additional children and three additional adults participated but were excluded from the final analysis for one of three reasons: (i) because other people were present during the experiment and prompted the subject with answers (two children, one adult), (ii) because they failed to use agreeing forms on a majority of the items (four children), or (iii) because they failed to classify eight out of ten very frequent words correctly (fourteen children, two adults). Reason (iii) 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 because 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 is 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 does not 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. The fact that so many children were excluded for this reason does not mean that our data is untrustworthy, only that different participants used different strategies for completing the task. In our analysis, we focus only on those participants whose responses were potentially informative about the acquisition of noun classes. The different strategies that children employed in completing our task undoubtedly raises an interesting methodological puzzle, but this puzzle 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 (e.g. a nonce word with semantic feature 'female', or a 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.

For every set of words with a given feature or set of features, the proportion of words assigned to each class was calculated. By comparing the differences between distributions for each cue type, we could determine which cues caused the distributions to change, and to what degree. This meant that for each set of words we had a distribution of noun class assignment for each age group.

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 refer to the proportion of words of each item type that was 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 refer to the proportion of the words that was 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 when it is absent.

In order to determine whether distributions really were different from one another, the Jensen-Shannon (JS) divergence was calculated between each relevant pairing of distributions (i.e. all of 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 1.

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

where $$M = \frac{P + Q}{2}$$

and $$D_{KL}(P \| M) = \sum P(i) \log\left(\frac{P(i)}{M(i)}\right)$$

This resulted in a distribution of possible JS divergences for the data under consideration. The JS divergence between two sets of interest (i.e. adults' use of a phonological cue for class 3 versus 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. We chose 10% as the cut-off criterion because this lay well outside the standard deviation of the mean of the measured set of JS divergences (mean = 0.15, *SD* = 0.11). The data was analyzed in this way instead of through *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 assigned to a given class, but also how the distribution was skewed with the introduction of a given cue, information that cannot be assessed with traditional hypothesis-testing statistics.

The comparison across groups that follows does not directly reference the JS divergences for a given cue, class, and group. Instead, it compares 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 of 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. Thus, the differences in the proportions presented below reflect actual differences in the classification of nouns by speakers in the experiment. What follows is a summary of the main findings from comparing these distributions. A full presentation of every item type and age group can be found in Appendix B. The following patterns emerged from this analysis.

- Classification of nonce words with phonological or semantic cues for classes 1, 2, and 3 reliably differed from classification of nonce words with no cues, but this classification did not differ across groups.
- Classification of nonce words with conflicting cues differed from classifications of words with only phonological or semantic cues for both child groups, but not for the adult group.

- Classification of real words with conflicting cues differed from the actual classification of these words for only the group of younger children.
- 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.

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 12).

	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

Table 12. Percentage of real words of each type correctly assigned to actual class. The * indicates that the JS divergence between the classification distributions of words with cues conflicting with actual class assignment and words in this class without conflicting cues was in the top 10% of the distribution of all JS divergences for real words (more than one standard deviation from the mean).

However, there are several things to point out in this data. First of all, in no case was classification perfect. The overall high percentages reveal that all three populations knew the words in question, and scores below 100% most likely reflect noise from the experimental task, rather than an imperfection in the classification of speakers as a group. The misclassification is distributed across lexical items, suggesting that this pattern stems from experimental noise, rather than lexical variation. It is important to remember that none of these groups approximates the typical sample populations used for psycholinguistic experiments carried out at research universities in the developed world. Even the adult speakers are not familiar with structured tests and games. Additionally, even the best testing conditions are less than ideal, as quiet rooms with no distractions were not available. Thus the less-than-perfect accuracy most likely reflects only the less-than-perfect conditions under which the experiment was carried out.

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. When cues conflicted with the actual class of the words, however, 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 a different class from that of which the word was a member. 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.

Classification of nonce words without cues. Next we consider the classification of nonce words with no predictive features. It must be noted, however, that the lack of predictive features 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 probability that any noun would be in any class. The results of this classification task are seen in Figure 2.

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

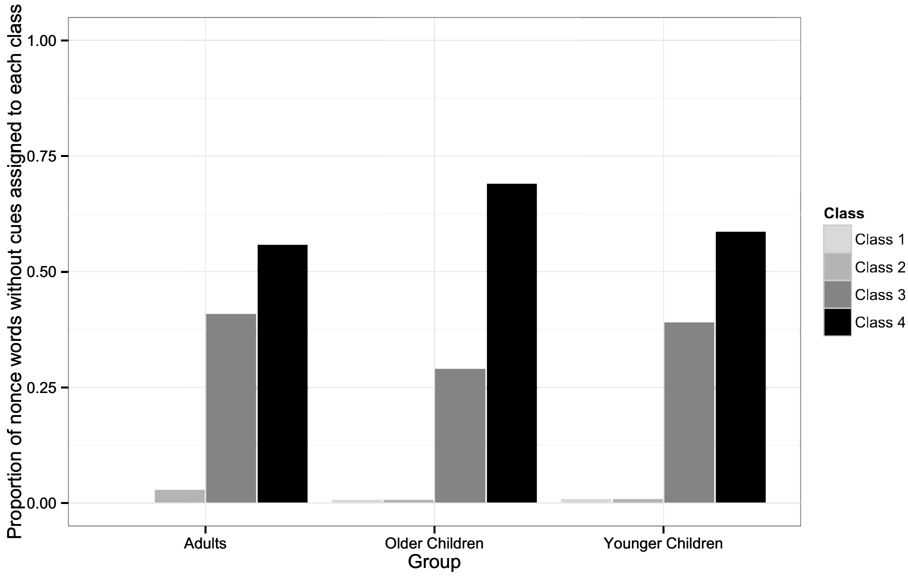

FIGURE 2. Classification of nonce words without cues: percentage of words assigned to each class by age group.

we look at the type frequencies of noun classes in the dictionary and type and token frequencies of noun classes in the corpus.

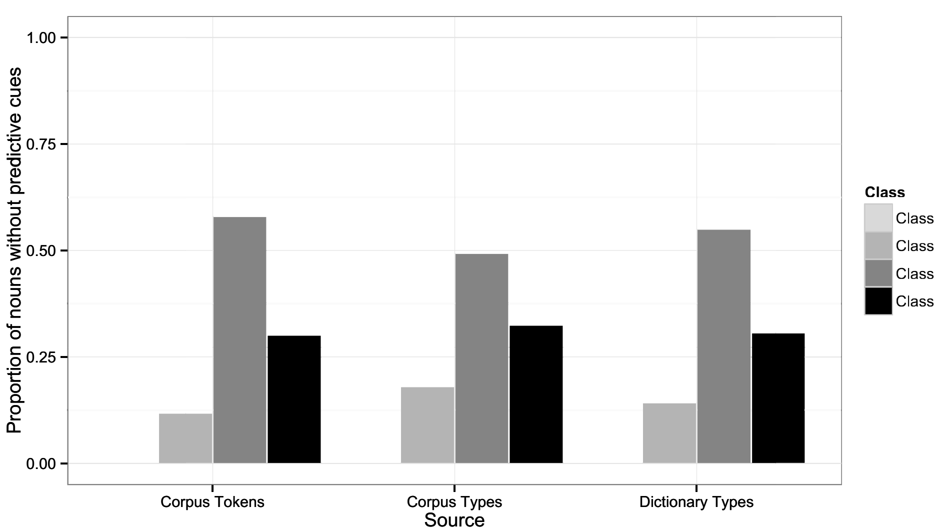

FIGURE 3. Frequencies of nouns without predictive cues in the dictionary and corpus.

While the default classification distribution seen in Fig. 2 does not precisely map onto any of those in Fig. 3, 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 concurrent modeling work addresses this issue (Gagliardi et al. 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 distribution 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 see below, these cues only work to skew this distribution in the direction indicated by the predictiveness of the cue; they do not work as rules assigning nouns to classes.

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 to 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. We call the class most strongly predicted by the noun-internal features the target class. In Table 13 we can see the proportion of words correctly assigned to the target class (the proportion of words classified according to the statistically strongest feature).

	BIOLOGICAL SEMANTIC	OTHER SEMANTIC	PHONOLOGICAL	CONFLICTING
YOUNG CHILDREN	54	8	61	**38***
OLDER CHILDREN	65	9	63	53
ADULTS	53	**23***	61	55

Table 13. Percentage of nonce words of each type correctly assigned to target class. The * indicates that the JS divergence between the classification distributions of words with these cue types and words with other cues to these classes was in the top 10% of the distribution of all JS divergences for nonce words (more than one standard deviation from the mean).

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 (i.e. the distribution reflected in Fig. 2 above). That is, while classification never approached 100%, it did, in some conditions outlined below, vary greatly from classification when no cue was present. 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 also very different from the default distribution. While the difference is not as extreme for classes 3 and 4, where a majority of the words ended up in the default distribution, when the relevant cues are present examination of the data by class shows that the vast majority of words end up there, many more so 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 tested only for class 2. Children do not appear to use this information at all, since the 8% and 9% of nonce words assigned to class 2 with the cues do not significantly differ from the 1% of cueless words assigned to class 2 (the JS divergence between these distributions does not fall in the top 10% of all JS divergences). For adults, by contrast, while the 23% of words with the other semantic cues 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, since 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 (for conditional probabilities of class given each feature, refer to Table 8 in §5.1 above). Thus, the class of 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 of 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 a cue 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 both by the noun-internal distributional cues (or lack thereof) and by a baseline distribution of nouns into classes.

6. General discussion. We found that speech to children contains ample evidence about the role of both noun-external and noun-internal cues in the assignment of nouns to noun classes. Moreover, a computational model of the noun-internal information allowed us to quantify the reliability of particular features in noun classification. Together, the corpus work and modeling work enabled us to ask what information learners use in acquisition, and whether they use this information in proportion to its statistical reliability.

The hypothesis that children use only the highly predictive noun-external distributional information to acquire noun classes predicted that children would have access to statistical regularities of inherent noun properties only late in the acquisition of noun classes, but that when they did, their generalizations should then mirror the adult ones. All of the hypotheses that noun-external information is not sufficient predicted that noun-internal information might not be used in proportion to its distribution in the input. In particular, the hypothesis that all noun-internal information is used (as opposed to only common semantic information or only phonological information) predicted that learners 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, since 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 toward the idea that children employ all available information in acquiring noun class for the following reasons.

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 an algorithm for building decision trees, the biological semantic features can classify with 100% accuracy, whereas the phonological features are massively less predictive. Nonetheless, children weigh the phonological cues more heavily when determining the class of a novel noun. This highlights a distinction between information that is present in the input and information that is used by learners, that is, the intake, in building the grammar of noun classification. 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 toward the utility of noun-internal distributional information in very early acquisition. First, across development, 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, even after a child is learning word meanings, 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 either from the early observation that phonological information is more useful, or from an a priori bias to prefer morphophonological information over semantic information when learning morphological dependencies.

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 the emerging 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. For now, it suffices to say that 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. Simply put, the information they use, the intake, does not match the information that is available in the input.

6.1. Specifying the role of the predictive noun-internal distributional information. Computational modeling allows 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 not only based on the predictive cues that a noun has, but also 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 get predictions for how each word type is classified by such a system, and can compare these (and thus our model) to actual classification, gaining a better understanding of the kind of categorization system this predictive information is playing a role in. Additionally, we are 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. For fuller discussion, see Gagliardi et al. 2012.

6.2. Mechanisms and further thoughts. Although this work points toward the hypothesis that children make use of both noun-external and noun-internal distributional properties to acquire noun classes, it has not addressed the precise mechanism that would require these two types of information in conjunction. There are two ways in which we are currently investigating exactly what the properties of this mechanism might be.

First, 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 are also important in investigating this mechanism further (see Gagliardi et al. 2012). By building explicit models of the acquisition process, we can 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 also allows 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 gives 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 make up only 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, does not 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 vast 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 and colleagues' 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. The architecture of this system, however, misses the overarching class structure: nouns with a given feature do not 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 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 to be a good fit for our data on noun classification, it is possible that our hypotheses about noun classification (e.g. that learners acquire classes by looking at properties of the items within a class in addition to the distribution of morphology across classes) might be capable of capturing irregular verb classes, and this topic deserves future investigation.

6.4. CONCLUDING REMARKS. In this article 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 makes use of in constructing a grammar. We have investigated the predictions that two hypotheses regarding this acquisition make about later behavior in noun classification. In doing so we have been able to draw inferences about 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 on regularities among nouns in a class. It also allows us to understand more about the kind of mechanism that 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. Overall, this work adds to our knowledge of language acquisition by contributing data from the acquisition of an understudied language, and, perhaps more importantly, highlights a domain where we can precisely measure the asymmetry between the child's input and the information used by the child in constructing a grammar.

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	riłu
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	ɣała
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

(TABLE A1. *Continues*)

WORD TYPE	ENGLISH	TSEZ
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	beło
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, phonological cue -i	novel food	tali
Nonce, 4, phonological cue -i	novel object	joni
Nonce, 4, phonological cue -i	novel object	q’omi
Nonce, 4, phonological cue r-	novel object	rega
Nonce, 4, phonological cue r-	novel food	ruło
Nonce, 4, phonological 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’	ɣuł
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’ƚ’ɣu
Real, 3, conflicting cues	‘sea’	raład

(TABLE A1. *Continues*)

WORD TYPE	ENGLISH	TSEZ
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, phonological cue -i	‘water’	łi
Real, 4, phonological cue -i	‘porridge’	qiqi
Real, 4, phonological cue -i	‘window’	aki
Real, 4, phonological cue r-	‘hand’	rełʼa
Real, 4, phonological cue r-	‘butter’	rił
Real, 4, phonological cue r-	‘key’	reka
Real, 4, agreeing cues	‘trash’	rešoni
Real, 4, agreeing cues	‘cradle’	rikini

TABLE A1. Full list of nouns used in classification experiment.

APPENDIX B: FULL CLASSIFICATION RESULTS FOR EACH ITEM TYPE

CLASSIFICATION OF REAL WORDS.

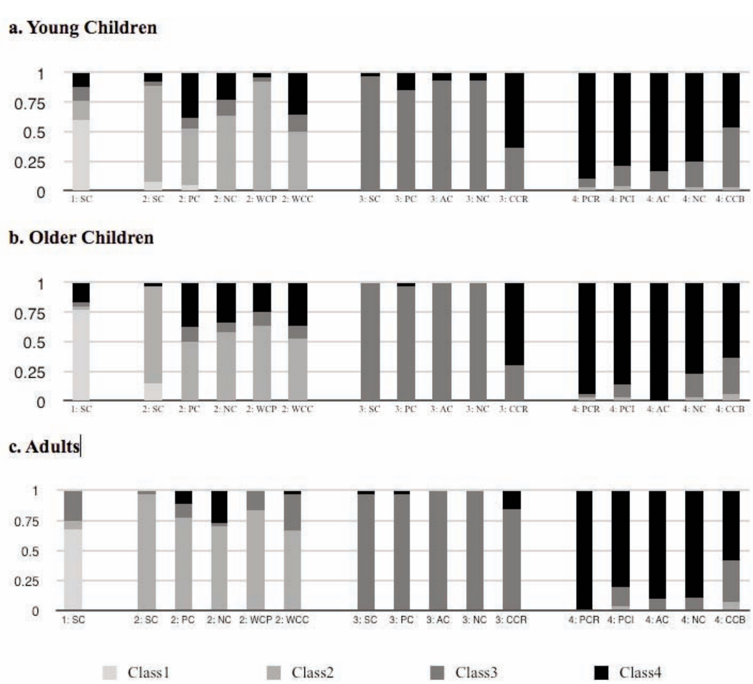

FIGURE B1. Each bar in the figure corresponds to a set of test items, grouped above by target class. The colors in the bars correspond to the proportion of words from this set assigned to the target class. Speakers generally assign nouns to the class they belong in, though when predictive information 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.

CLASSIFICATION OF NONCE WORDS.

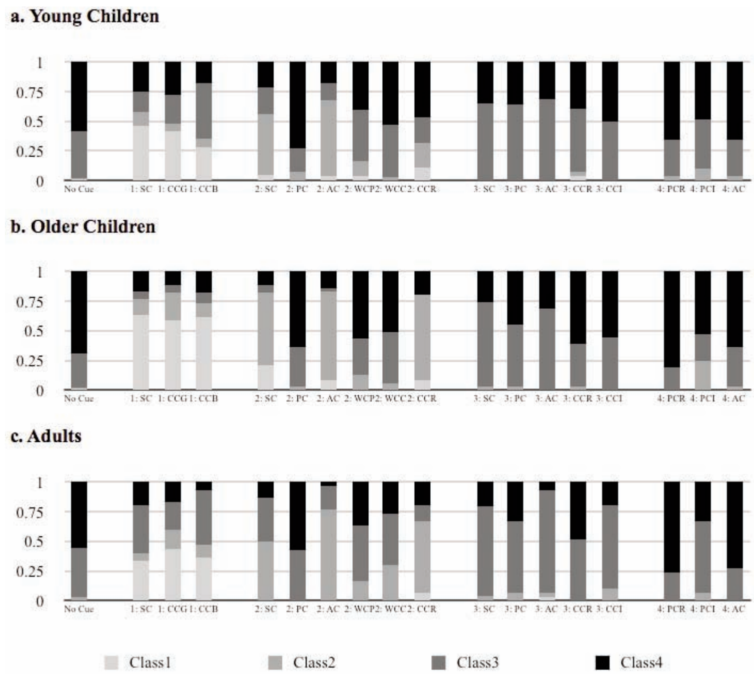

FIGURE B2. While nonce words show more noise, there is an evident effect of biological semantic cues on all groups, though only adults appear to use other semantic cues. Phonological cues are used, except those for class 2 (probably related to a misrepresentation of the frequency of this cue in the input). When semantic and phonological information conflict, children appear most likely to use phonological information and adults semantic (except for when this information is the nonworking phonological cue for class 2).

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

TABLE B1. Codes for item types in Figures B1–B2.

REFERENCES

Ambridge, Ben; Julian M. Pine; Caroline F. Rowland; Rebecca L. Jones; and Victoria Clark. 2009. A semantics-based approach to the 'no negative evidence' problem. *Cognitive Science* 33.1301–16.

Bokarev, Evgeny A. 1967. *Cezskie (didojskie) jazyki Dagestana*. Moscow: Nauka.

Booth, Taylor L., and Richard A. Thompson. 1973. Applying probability measures to abstract languages. *IEEE Transactions on Computers* 22.5.442–50.

Braine, Martin D. S. 1987. What is learned in acquiring word classes—A step toward an acquisition theory. *Mechanisms of language acquisition*, ed. by Brian MacWhinney, 65–87. Hillsdale, NJ: Lawrence Erlbaum.

Chomsky, Noam. 1959. Review of *Verbal behavior*, by B. F. Skinner. *Language* 35.26–58.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Comrie, Bernard, and Maria Polinsky. 1998. The great Daghestanian case hoax. *Case, typology and grammar: In honor of Barry J. Blake*, ed. by Anna Siewierska and Jae Jung Song, 95–114. Amsterdam: John Benjamins.

Comrie, Bernard, and Maria Polinsky. 1999. Some observations on class categorization in Tsez. *Studies in Caucasian linguistics: Selected papers of the Eighth Caucasian Colloquium*, ed. by Helma van den Berg, 125–39. Leiden: Universiteit Leiden.

Comrie, Bernard; Maria Polinsky; and Ramazan Rajabov. 1998. Tsezian languages. Leipzig: Max Planck Institute for Evolutionary Anthropology, ms.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Crain, Stephen. 1991. Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 4.597–650.

Elman, Jeffrey L.; Elizabeth A. Bates; Mark H. Johnson; Annette Karmiloff-Smith; Domenico Parisi; and Kim Plunkett. 1996. *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Fodor, Janet D. 1998. Parsing to learn. *Journal of Psycholinguistic Research* 27.339–74.

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

Gagliardi, Annie; Naomi H. Feldman; and Jeffrey Lidz. 2012. When suboptimal behavior is optimal and why: Modeling the acquisition of noun classes in Tsez. *Proceedings of the 34th annual conference of the Cognitive Science Society*, 360–65.

Gerken, LouAnn; Rachel Wilson; Rebecca L. Gomez; and Erica Nurmsoo. 2002. Linguistic category induction without reference: The importance of correlated cues. Tucson: University of Arizona, ms.

Gerken, LouAnn; Rachel Wilson; and William Lewis. 2005. Infants can use distributional cues to form syntactic categories. *Journal of Child Language* 32.249–68.

Gold, E. Mark. 1967. Language identification in the limit. *Information and Control* 10. 447–74.

Halle, Morris, and K. P. Mohanan. 1985. Segmental phonology of Modern English. *Linguistic Inquiry* 16.1.57–116.

Harris, Zellig. 1951. *Methods in structural linguistics*. Chicago: University of Chicago Press.

Hay, Jennifer B., and R. Harald Baayen. 2005. Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences* 9.342–48.

Hu, Qian. 1993. *The acquisition of Chinese classifiers by young Mandarin-speaking children*. Boston: Boston University dissertation.

Jackendoff, Ray. 2002. *Foundations of language*. Oxford: Oxford University Press.

Jelinek, Frederick. 1990. Self-organized language modeling for speech recognition. *Readings in speech recognition*, ed. by Alex Waibel and Kai-Fu Lee, 450–506. San Mateo: Morgan Kaufmann.

Karmiloff-Smith, Annette. 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.

Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45.715–62.

Lewis, John D., and Jeffrey L. Elman. 2001. A connectionist investigation of linguistic arguments from the poverty of the stimulus: Learning the unlearnable. *Proceedings of the 23rd annual conference of the Cognitive Science Society*, 552–57.

Lin, James. 1991. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* 371.145–51.
MacWhinney, Brian. 1978. The acquisition of morphophonology. *Monographs of the Society for Research in Child Development* 43.1/2.1–123.
Mills, Anne E. 1985. The acquisition of German. *The crosslinguistic study of language acquisition*, vol. 1, ed. by Daniel I. Slobin, 141–254. Hillsdale, NJ: Lawrence Erlbaum.
Mills, Anne E. 1986. *The acquisition of gender: A study of English and German*. Berlin: Springer.
Pérez-Pereira, Miguel. 1991. The acquisition of gender: What Spanish children tell us. *Journal of Child Language* 183.571–90.
Pearl, Lisa. 2011. When unbiased probabilistic learning is not enough: Acquiring a parametric system of metrical phonology. *Language Acquisition* 18.2.87–120.
Pearl, Lisa, and Jeffrey Lidz. 2009. When domain-general learning fails and when it succeeds: Identifying the contribution of domain specificity. *Language Learning and Development* 5.4.235–65.
Pearl, Lisa, and Amy Weinberg. 2007. Input filtering in syntactic acquisition: Answers from language change modeling. *Language Learning and Development* 3.1.43–72.
Pinker, Stephen. 1979. Formal models of language learning. *Cognition* 7.217–83.
Pinker, Stephen. 1984. *Language learnability and language development*. Cambridge, MA: Harvard University Press.
Pinker, Stephen. 1991. Rules of language. *Science* 253.530–35.
Pinker, Stephen, and Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28.73–193.
Plaster, Keith; Maria Polinsky; and Boris Harizanov. 2013. Noun classes grow on trees: Noun classification in the North-East Caucasus. *Language typology and historical contingency: In honor of Johanna Nichols*, ed. by Balthasar Bickel, Lenore A. Grenoble, David A. Peterson, and Alan Timberlake, 153–70. Amsterdam: John Benjamins.
Polinsky, Maria. 2000. Tsez beginnings. *Berkeley Linguistics Society* 25.14–29.
Polinsky, Maria, and Dan Jackson. 1999. Noun classes: Language change and learning. *Cognition and function in language*, ed. by Barbara A. Fox, Dan Jurafsky, and Laura A. Michaelis, 29–58. Chicago: University of Chicago Press.
Quinlan, J. R. 1993. *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann.
Rodina, Yulia. 2009. *Semantics and morphology: The acquisition of grammatical gender in Russian*. Tromsø: University of Tromsø dissertation.
Rumelhart, David E., and James L. McClelland. 1986. On learning the past tenses of English verbs. *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 2, ed. by James L. McClelland, David E. Rumelhart, and the PDP Research Group, 216–71. Cambridge, MA: MIT Press.
Sankoff, David. 1971. Branching processes with terminal types: Application to context-free grammars. *Journal of Applied Probability* 8.233–40.
Tucker, G. Richard; Wallace E. Lambert; and Andre Rigault. 1977. *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. The Hague: Mouton.
Valian, Virginia. 1999. Input and language acquisition. *Handbook of child language acquisition*, ed. by William C. Ritchie and Tej K. Bhatia, 497–530. New York: Academic Press.
Witten, Ian, and Eibe Frank. 2005. *Data mining: Practical machine learning tools and techniques*. 2nd edn. San Francisco: Morgan Kaufmann.
Yang, Charles D. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.
Yang, Charles D. 2004. Universal grammar, statistics or both? *Trends in Cognitive Sciences* 8.451–56.

Gagliardi
Department of Linguistics
Harvard University
Boylston Hall, 3rd Floor
Cambridge, MA 02138
[gagliardi@fas.harvard.edu]

[Received 2 December 2010;
revision invited 5 July 2011;
revision received 28 October 2011;
revision invited 16 July 2012;
revision received 26 October 2012;
accepted 17 April 2013]

Lidz
Department of Linguistics
University of Maryland, College Park
1401 Marie Mount Hall
University of Maryland
College Park, MD 20742-7505
[jlidz@umd.edu]