In Appendix A we provide further explanation of how specific situation definitions allow for more efficient grammar encoding. We emphasise that we are not tied to this particular grammar description. It is only one example of many possible grammatical descriptions that can be used within our framework. In appendix B we show an example MDL calculation for the restriction on *is* contraction. We explicitly write out the relevant portion of the original and new grammars and show how to use MDL to evaluate learnability of this restriction given language data estimated from the CHILDES corpus. In Appendix C we show the relevant portions of new vs. original grammars for all the constructions we analyze.

Appendix A: Approximation of speaker's grammar

Specific-situation rules allow for a more efficient encoding of the language data by specifying more accurate rule probabilities for specific situations. Coding theory dictates that the most efficient representation will occur when an element n is encoded using approximately $-\log_2(p_n)$ bits, where p_n is the probability of the nth element in data. Specific-situation rules will specify a list of situation specific syntax-rules followed by their situation-dependent probabilities, which will be used to implement a more efficient, situation-dependent code. The notion that speakers have knowledge of local probabilities for different variations of particular utterances is reasonable as experimental studies have shown that these probabilities are known relative to local specific linguistic situations (Bresnan, 2006; Jurafsky, 2003).

Specific-situation rules thus provide efficient ways of representing situations such as contractions. Here we use the specific-situation rules to specify which pairs of words are contractable and how often contraction occurs. An alternative to specific-situation rules for contractions would be to use a special symbol for encoding [contract] every time a pair of words is contracted in language. As most word pairs in language are not contracted, the use of a [contract] symbol in the language encoding would be extremely costly (require a long code length) as this symbol would occur rarely. Instead it is much more efficient to specify all contract-able word pairs (e.g. words followed by *is* or word pairs, *want to* and *going to*). Then, when encoding language, all contract-able word pairs will be followed by symbols specifying either [contract] or [no-contract]. These will be encoded according to the contraction probability of the specific contract-able word pair, which is much higher than that of general word pairs, resulting in much shorter encoding lengths. Using individual specific-situation rules that specify contraction probabilities unique for each contract-able word pair provides code-length savings. To illustrate this, let us imagine that in general, contractions occur only 50% of the time over all contract-able word pairs used in language. If contraction rules were used generally, every presence of a contraction would be encoded with $-\log_2(0.5)=1$ bit. Alternatively, a specific-situation rule could specify different contraction probabilities for different word pairs. In particular, let's consider the specific-situation rule about a word pair that contains *is*. This specific-situation rule would take into account the fact that contraction of *is* occurs 85% of the time and thus encode the *is* contraction with $-\log_2(0.85)=0.23$ bits This produces a savings of 0.77 bits per encoding of the *is* contraction. Thus, the addition of a simple specific-situation rule can result in a much more efficient encoding of language.

Appendix B: Example Calculation of MDL applied to a corpus:

In order to clearly describe our methodology we will explain our calculations using the restriction on *is* contraction as an example. First, the portions of original and new grammars relevant to the linguistic construction being considered are explicitly written out (see below). Everything enclosed in brackets is counted as one symbol, and these symbols all have semantic meanings that are either defined under specific situation definitions or assumed to be known by the language learner. For example, it is assumed that the learner knows that the symbol [situation definition] indicates that specific situation symbols will be defined below. Between [situation definition] and [end] symbols, we assume that the learner knows that concept symbols, which include specific situation symbols (i.e. [something is]), are always followed by their definition. Each definition is ended by an end symbol, #. Unlike the main body of specific-situation rules (introduced by the [situation] symbol), situation definitions (introduced by the [situation definition] symbol) do not have encoding probabilities. While in this present situation of *is* contraction, we only need to define specific situation symbols, in other examples, it will be necessary to define further concept symbols which will help in defining the specific situation symbol. Below, we assume that the learner understands the concept [something], to mean any word that appears before *is*. Similarly, we assume that the learner understands [punctuation] to be any punctuation such as a period, comma, question mark etc. Again, as mentioned in the text, the learner does not need to understand the formal grammar of punctuation per se; they need only recognize the appropriate phonological string such as particular word intonations followed by a pause. We also assume that all specific-situation rules begin with a [situation] symbol. Following this is a list of symbols for each of the specific situation (i.e [contract]) followed by the new occurrence probabilities for the situation specified. An [end] symbol concludes the specific situation rules.

Approximating grammar values

A calculation of grammar length costs requires knowledge of the usage-frequency of each symbol type, $f_{s,orig}$, and the total number of symbol $F_{total,orig}$. These respective values in the new grammar, $f_{s,new}$ and $F_{total,new}$ can be calculated directly from the differences between new vs. original grammatical descriptions that are explicitly written out. We note that we only need to evaluate values of $f_{s,orig}$ for the symbols that differ between the grammars that we are comparing, as the values of $f_{s,orig}$ that remain the same between original and new grammars do not affect grammar length differences (see main text for further explanation). Some of these symbols will occur only in the relevant portion of our specific-situation rules. Other symbols will also occur in the rest of the grammar that we have not specified (either in other specific-situation rules or in grammar or vocabulary rules. Examples of symbols that also appear in the rest of the grammar include: the end symbol, #, the symbols [contract] and [no-contract]. The number of end symbols, #, in the original grammar (not including the relevant specified portion) we estimate to be $\frac{F_{total,orig}}{5}$, which means that we assume, for the entire grammar on average, one out of five symbols will be a end symbol. The symbols [contract] and [no-contract] we will estimate to have occurred about 12 times each in the specific-situation rules. The estimated number of 12 comes from the assumed knowledge for contractions of not, going to, want to, got to, and 8 auxiliary verbs,). An approximation is also needed for $N_{situation,orig}$, the number of specific situation definitions already present in the original grammar. This determines the number of [situation] and [end] symbols in the grammars. Here we use a conservative estimate $f_{Nsituation,orig}=0$. Note that the lower the estimate of $f_{s,orig}$, the greater the grammar cost (the opposite of that for $F_{total,orig}$). Thus estimating $f_{Nsituation,orig}$ =0 results in an upper bound estimate for grammar length differences, i.e. greater difficulty of learning a new linguistic rule. Similarly, for all other symbols we will also assume that each symbol appears for the first time in the specific portion of the grammars that we consider. We then need to estimate the value for $F_{total,orig}$. Here, the following key principle must be kept in mind: The larger the values for $F_{total,orig\ ig}$, the larger the grammar cost, as well as the larger the additional cost of acquiring new rules. Thus when estimating these values, erring on the large side for $F_{total,orig}$ will result in *upper bound* estimates for grammar costs. It is also important to keep in mind that because of the logarithmic nature of encoding, a change in the order of magnitude for $F_{total,orig}$ will result only in small differences in grammar lengths. For example, consider a symbol x that is used only once in the grammar. If $F_{total,orig}$ were estimated to be 200, the symbol x would cost $-\log_2(1/200) = 7.6$ bits to encode. If $F_{total,orig}$ were estimated to be 2000, the symbol x would cost $-\log_2(1/2000) = 11$ bits to encode. We evaluated results using $F_{total,orig}$=200 and $F_{total,orig}$=100000 and found that they did not differ qualitatively.

Example MDL calculation on contraction of 'is'
The original and new grammars for the relevant grammar portion of *is contraction* are written out below. Relative occurrence probabilities of contraction are estimations from CHILDES corpus.

Original grammar:
[situation definition]
    [something is] [something] is #
[end]

[situation] [something is]
    [contract] 0.88
    [no-contract] 0.12
[end]

New grammar:
[situation definition]
    [something is] [something] is #
    [contractable is] [something is] [not-punctuation] #
    [not-contractable is] [something is] [punctuation] #
[end]

[situation] [contractable is]
    [contract] 0.85
    [no-contract] 0.15
[end]

[situation] [not-contractable is]
    [no-contract] 1
[end]

Evaluation of encoding length for new vs. original grammar

Equation 3 from the main paper can then be used to calculate coding length differences between the new and original grammars. For convenience we repeat equation 3 here:

$$\Delta_{grammar} = -\sum_{s,relevant}\left(f_{s,new}\log_2\frac{f_{s,new}}{F_{total,new}} - f_{s,orig}\log_2\frac{f_{s,orig}}{F_{total,orig}} - f_{s,orig}\log_2\frac{F_{total,orig}}{F_{total,new}} \right) - F_{total,orig}\log_2\frac{F_{total,orig}}{F_{total,new}} + (\Delta N_{rules} + \Delta N_{symbols})C_{prob}$$

(3)

Where *s, relevant* are the symbols that differ between original and new grammars, $f_{s,new}$ and $f_{s,orig}$ are occurrence frequencies of symbol *s* in new and original grammars, $F_{total,new}$ and $F_{total,orig}$ are total symbol frequencies in new and original grammars, ΔN_{rules} *and* $\Delta N_{symbols}$ are the differences in the number of rules and symbols respectively between new vs. original grammars. C_{prob} is assumed to be = 6.6 bits which allows probabilities to be encoded to a two decimal accuracy (as explained in the main text).

Grammar length contributions are calculated by making a list of the occurrence frequencies for symbols that differ between original and new grammars as follows:

Table B1: *Original grammar symbol summary*

Symbol list	# occurrences in grammar	contribution to grammar length
[situation definition]	1	$\log_2(1/F_{total,orig})$
[something is]	2	$2 \times \log_2(2/F_{total,orig})$
[something]	1	$\log_2(1/F_{total,orig})$
is	1	$\log_2(1/F_{total,orig})$
[end]	2	$2 \times \log_2(2/F_{total,orig})$
[situation]	1	$\log_2(1/F_{total,orig})$
[contract]	12	$12 \times \log_2(12/F_{total,orig})$
[no-contract]	12	$12 \times \log_2(12/F_{total,orig})$
#	$N_{\#,orig}$	$N_{\#,orig} \times \log_2(N_{\#,orig}/F_{total,orig})$

We also count the number of symbols types and the number of rules.
$N_{symbols,orig}$= 9, $N_{rules,orig}$=2

Here $N_{symbols,orig}$ and $N_{rule,origs}$ are only counted from the portion of the grammar description that will change between original vs. new grammars because eventually only the difference is required. Also note that the occurrence frequencies of [contract] and [no-contract] are assumed to equal 12 in the original grammar, as explained in the above section *Approximating grammar values*. Assuming $F_{total,orig}$ =200, $N_{\#,orig}$= $F_{total,orig}$ /5+1, (as described in the above) we can evaluate the following term from equation 3:

$$\sum_{s,relevant} f_{s,orig}\log_2\frac{f_{s,orig}}{F_{total,orig}} = -248.3 \text{ bits.}$$

Table B2: *New grammar symbol summary*

Symbol list	# occurrences in grammar	contribution to grammar length
[situation definition]	1	$\log_2(1/F_{total,new})$
[contractable is]	2	$2 \times \log_2(2/F_{total,new})$
[not-contractable is]	2	$2 \times \log_2(2/F_{total,new})$
[something is]	3	$3 \times \log_2(3/F_{total,new})$
[something]	1	$1 \times \log_2(1/F_{total,new})$
is	1	$1 \times \log_2(1/F_{total,new})$
[not-punctuation]	1	$\log_2(1/F_{total,new})$
[punctuation]	1	$\log_2(1/F_{total,new})$
[end]	3	$3 \times \log_2(3/F_{total,new})$
[situation]	2	$2 \times \log_2(2/F_{total,new})$
[contract]	12	$12 \times \log_2(13/F_{total,new})$
[no-contract]	13	$13 \times \log_2(14/F_{total,new})$
#	$N_{\#,orig}+2$	$(N_{\#,orig}+2) \times \log_2(N_{\#,orig}+2)/F_{total,new})$

Again we also count the number of symbols types and the number of rules: $N_{symbols,new}= 13$, $N_{rules,new}=3$

For this new grammar we need make the additional calculation of $F_{total,new}= F_{total,orig}+\Delta F_{total}$, where ΔF_{total} is the difference between the total number of symbol occurrences in the new grammar vs. the original grammar. From the second columns of the contribution summaries from new and original grammars we have $\Delta F_{total} =12$ so $F_{total,new}=212$. Now can evaluate the following term from equation 3:

$$-\sum_{s,relevant} f_{s,new} \log_2 \frac{f_{s,new}}{F_{total,new}} = 316.9 \text{ bits.}$$

Additionally, we have to evaluate the following terms from equation 3:

$$(\Delta N_{rules} + \Delta N_{symbols})C_{prob} = 33 \text{ bits}$$

$$(-F_{total,orig} + \sum_{s,relevant} f_{s,orig})\log_2 \frac{F_{total,orig}}{F_{total,new}} = 10.68 \text{ bits}$$

Thus the difference between new and original grammars is:
316.9-248.3 + 33 + 10.67 bits. This means an ideal MDL language learner will have to accrue a language encoding savings of 112.3 bits under the new grammar to make the new grammar worth adopting over the original grammar.

Evaluation of language encoding lengths under new vs. original grammars

Under the new grammar, new probabilities will be assigned to particular linguistic rules. For our purposes, these probabilities are estimated from the occurrence frequencies of each linguistic context in the Brown and Bates corpus from CHILDES. In reality, a child does not have prior access to language probabilities. Instead, these probabiltiies are estimated and updated continuously as the child experiences language. Thus, an adaptive coding scheme which updates probability estimates may be more accurate. However, because the purpose of this paper is to get order of magnitude estimates of learnability, the precision offered by an adaptive probability estimation model is unnecessary. If one were to use an adaptive estimation strategy, grammar lengths in general would be longer to account for the updating probabilities.

In order to specify rule probabilities for the new grammar we also need to distinguish between un-contracted occurrences where contraction is allowed vs. not allowed. We also have to specify all the relevant linguistic situations, whose encoding costs will have changed, and use their relative occurrence frequencies in corpora to evaluate their costs under the new vs. original grammar. In the situation of *is* contraction, there are 3 relevant linguistic situations with results summarized in the table below: 1) *is* contracted, contraction allowed, (abbreviated as Contract Allowed) 2) *is* not contracted, contraction not allowed, (abbreviated as No-contract, Allowed) 3) *is* not contracted, contraction not allowed (abbreviated as No-contract, Not-allowed).

Table B3: Encoding analysis for *is* contraction

Situation	freq	original encoding length (bits)	new encoding length (bits)	Savings
1. Contract Allowed	10043	$-\log_2 \frac{10{,}043}{(1417+382)+10{,}043} = 0.24$	$-\log_2 \frac{10{,}043}{1417+10{,}043} = 0.2$	0.04
2. No-contract Allowed	1417	$-\log_2 \frac{(1417+382)}{(1417+382)+10{,}043} = 2.73$	$-\log_2 \frac{1417}{1417+10{,}043} = 3.0$	-0.27
3. No-contract Not-Allowed	382	$-\log_2 \frac{(1417+382)}{(1417+382)+10{,}043} = 2.73$	$-\log_2 (\frac{382}{382}) = 0$	2.73

Freq is the frequency for the specific linguistic situation occurring in the combined Brown and Bates corpora in CHILDES. *Original encoding length* and *new encoding length* are the encoding lengths under the original and new grammars, respectively. *Savings* is the number of bits saved under the new grammar vs. the original grammar every time the specific linguistic situation appears.

From corpora analysis, we find that situations 1,2, and 3 occur 10,043, 1417, and 382 times respectively. The calculation of normalization probabilities is relative to all other possibilities under a specific situation rule. Because the original grammar always allows contraction, it does not distinguish between situations 2 and 3, and they are treated as a single situation where contraction occurs (382+1417) times under the original grammar. Thus under the original grammar, the not-contracted encoding lengths are exactly the same for contraction allowed and not-allowed situations. The new grammar offers encoding savings for appearances of *is* where contraction occurs (0.04 bits) and where contraction does not occur and is not allowed (2.73 bits). There is an encoding loss (0.27 bits) for every appearance of *is* where contraction does not occur but is allowed. However, because situations 1 and 3, which afford savings, appear often enough relative to situation 2, which causes losses, the overall encoding length is smaller under this new grammar. (Note: a more specific grammar description does not necessarily result in shorter overall encoding lengths. Under MDL, a learner would only choose the grammars that do. These are the grammars that represent the language most accurately.) In order to calculate total encoding savings we assume that the relative occurrence frequencies estimated from corpora for all specific linguistic situations is representative of their average relative occurrence probabilities in language. We can then calculate the total savings per occurrence of the j^{th} specific linguistic situation, $TotalSavings_j$, by summing the frequency weighted encoding gains/losses over all situations and dividing by the frequency of occurrence for the j^{th} specific linguistic situation.

$$TotalSavings_j = \frac{\sum_i freq \times save_i}{freq_j}$$

In our results, we report encoding savings per occurrence of the linguistic restriction (specific linguistic situation 3 in this example). Thus we have

$$TotalSavings_3 = \frac{10{,}043 \times 0.04 + 1417 \times (-0.27) + 382 \times 2.78}{382} = 2.9 bits$$

Thus every time *is* occurs in the non-contractable situation (i.e. at the end of a sentence) 2.9 bits is saved if a language speaker uses the new grammar vs. the original grammar. From our calculation above we know that the new grammar description costs 112 bits more than the original grammar description. Thus savings of 112 bits is needed to make our new grammar worth while. So we can calculate the number of occurrences of not-contractable *is* will be $\frac{112 bits}{2.9 bits / occurrence} \approx 39 occurrences$.

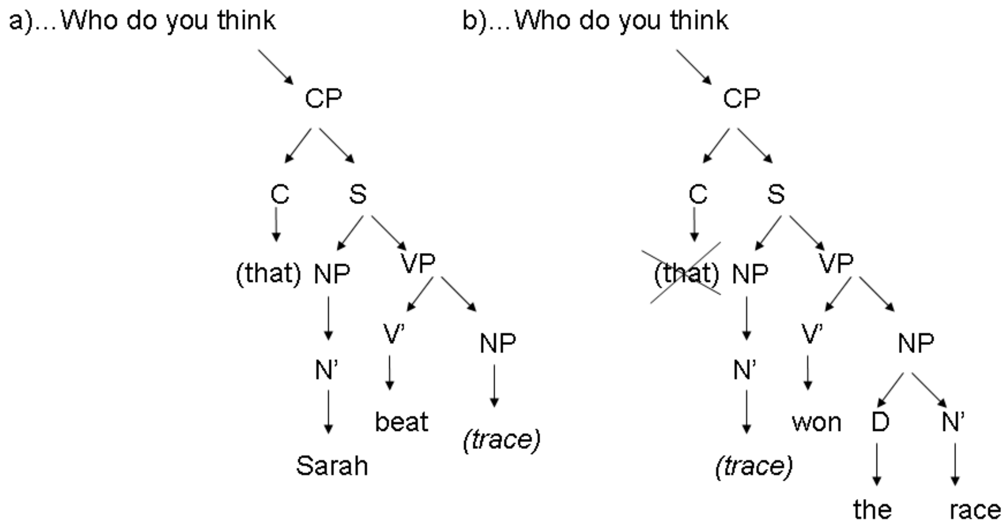
Appendix C: Specific-situation Grammars

Here our purpose is not to present linguistic theory for these constructions. Instead, these are only the grammar representations we chose for demonstrating our framework. Our framework is general and can be used under many different grammar formulations, which may produce opposing learnability results.

Specific-situation grammar: optionality of that

Here we assume the speaker knows about the optionality of the complementizer *that* in Wh-questions (i.e. *Who do you think that she called?*). This means understanding the trace left by the wh-word (e.g. *who, what*). We also assume that the speaker implicitly understands the tree-form phrase-structure of this sentence and understands the notions of non-trivial parent and sibling in the tree structure. Again, this knowledge does not have to be in the form of an explicit formal grammar, but knowledge of the equivalent patterns in phonological strings or prosodic cues. We feel this is a reasonable assumption because we are considering a speaker who already knows how to insert *that* into the clause of wh-questions. We use this knowledge to describe the concept of licensing in the situation definition of our new grammar: If sibling of A = the non trivial parent of B, then A licenses B. For example in *Who did you think Sarah beat*?, *(trace)* occupies the object position in the complement phrase and *that* licenses *Sarah* (Figure C1a). Alternatively, in *Who did you think (trace) won the race?*, (*trace)* occupies a subject position in the complement phrase phrase and *that* licenses *(trace)* (Figure C1b). The reduction of *that* is only optional when *that* does not license a trace, as in the first sentence. Otherwise, reduction is mandatory.

Figure C1: that-reduction and trace licensing



Original grammar:
[situation definition that-option]
    [that-option] CP-> C S #,
    [that-present] C -> that #
    [that-reduction] C-> [empty] #
[end]

[situation] [that-option]
    [that-present] prob
    [that-reduction] prob
[end]

New grammar:
[situation definition that]
    [that-option] CP-> C S #,
    [that-present] C -> that #
    [that-reduction] C-> [empty] #
    [licenses] if sibling A = [non-trivial parent] B then A [licenses] B #
    [that-reduction optional] C [licenses] [trace] #
    [that-reduction mandatory] C [licenses] [not-trace] #

[end]

[situation] [that-reduction optional]
    [that-present] prob
    [that-reduction] prob
[end]

[situation] [that-reduction mandatory]
    [that-reduction] prob

[end]

Specific-situation grammar: general contraction

Below we show the general structure of new vs. original grammars for the portions relevant to learning restrictions on contraction rules. This general structure is customized for each specific contraction situation by replacing word1 and word2 with the two words involved in the contraction. The definitions of contractable vs. not-contractable contexts specified within the situation definition of the new grammar are unique to each situation. These unique definitions are shown for individual constructions by showing their situation definition under the new grammar.

Original grammar:

```
[situation definition word1 word2]
        [word1 word2] word1 word2 #
[end]

[situation] [word1 word2]
        [contract]  prob
        [no-contract]  prob
[end]
```

New grammar:

```
[situation definition word1 word2]
        [word1 word2] word1 word2 #
        [contractable word1 word2] contractable_definition  #
        [not-contractable word1 word2] not_contractable_definition  #
[end]

[situation] [contractable word1 word2]
        [contract]  prob
        [no-contract]  prob
[end]

[situation] [not-contractable word1 word2]
        [no-contract]  prob
[end]
```

New grammar situation definition of: *want to contraction*

New grammar:

```
[situation definition want to]
        [want to] want to #
        [contractable want to] x wants x to #
        [not-contractable want to] x wants y to #
[end]
```

New grammar situation definition of: *going to contraction*

New grammar:
[situation definition going to]
    [going to] going to #
    [contractable going to] [going to] [verb] #
    [not-contractable going to] [going to] [a place] #
[end]

New grammar situation definition of: *is contraction*

New grammar:
[situation definition something is]
    [something is] [something] is #
    [contractable is] [something is] [not-punctuation] #
    [not-contractable is] [something is] [punctuation] #
[end]

New grammar situation definition of: *what is contraction*

New grammar:
[situation definition what is]
    [what is] what is #
    [contractable what is] [what is] [not it punctuation] #
    [not-contractable what is] [what is] [it punctuation] #
[end]

New grammar situation definition of: *who is contraction*

New grammar:
[situation definition who is]
    [who is] who is #
    [contractable who is] [who is] [not it punctuation] #
    [not-contractable want to] [who is] [it punctuation] #
[end]

Specific-situation grammar: general dative alternation

Below we show the general structure of new vs. original grammars for the portions relevant to learning restrictions on the dative alternation. This general structure is customized for each specific dative alternation situation by replacing verb1 with the more common word that does undergo the alternation and verb2 with the semantically similar verb that is overgeneralized in the original grammar to be allowed to alternate like verb1. The verbs whose dative alternation restrictions that we choose to analyze along with the dative alternating verbs they are associated with are listed in the main paper.

Original grammar:

[situation definition verb1/verb2]

[direct-dative] VP->V NP NP #

[prepositional-dative] VP->V NP PP #

[dative-alternation verb1/verb2] verb1 verb2 #

[end]

[situation] [dative-alternation verb1/verb2]

[direct-dative] prob

[prepositional-dative] prob

[end]

New grammar:

[situation definition verb1/verb2]

[direct-dative] VP->V NP NP #

[prepositional-dative] VP->V NP PP #

[dative-alternation verb1] verb1 #

[prepositional-only verb2] verb2 #

[end]

[situation] [dative-alternation verb1]

[direct-dative] prob

[prepositional-dative] prob

[end]

[situation] [prepositional-only verb2]

[prepositional-dative] prob

[end]

Specific-situation grammar: general restrictions on transitivity

Below we show the general structure of new vs. original grammars for the portions relevant to learning restrictions on verb transitivity. This general structure is customized for each specific transitivity situation by replacing verb1 with the more common word that does is ambitransitive and verb2 with the semantically similar verb that can only be transitive (or intransitive), but which is assumed in the original grammar to be ambitransitive like verb1. The verbs whose transitivity that we choose to analyze along with the ambitransitive verbs they are associated with are listed in the main paper. Treatment of intransitive-only verbs is the same as that of transitive-only verbs. Here, the concept of [transitive] is simply replaced by [intransitive], as shown in parentheses below.

Original grammar:

[situation definition verb1/verb2]

[transitive] VP->V NP #

[intransitive] VP-> V’ #

[ambitransitive verb1/verb2] verb1 verb2 #

[end]

[situation] [ambitransitive verb1/verb2]
    [transitive] prob
    [intransitive] prob
[end]

New grammar:
[situation definition verb1/verb2]
    [transitive] VP->V NP #
    [intransitive] VP-> V' #
    [ambitransitive verb1] verb1 #
    [transitive (intransitive) only verb2] verb2 #
[end]

[situation] [ambitransitive verb1]
    [transitive] prob
    [intransitive] prob
[end]

[situation] [transitive (intransitive) only verb2]
    [transitive] prob
    ([intransitive] prob)
[end]

Reference List

Bresnan, J. (2006). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. *Proceedings of the International Conference on Linguistic Evidence,* 75-96.

Jurafsky, D. (2003). Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production. In R.Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (Cambridge, MA: MIT Press.