Where do syllables come from?

Evelyn Martens and Walter Daelemans and Steven Gillis and Helena Taelman
Universitaire Instelling Antwerpen
Universiteitsplein 1
2610 Wilrijk
Belgium

Abstract

Young children are able to segment words into syllables, even though there are no perceptual or acoustic cues that indicate syllable boundaries in the primary linguistic data. We show that information about word boundaries can be used to predict syllable boundaries by replicating the results of experiments done by Gillis and De Schutter (1996) with children who syllabified Dutch disyllabic monomorphemes with a single intervocalic consonant. Word boundary probabilities were statistically computed in child language corpora and used to predict syllable boundaries with a simple statistical model. The children's syllabification behavior could be simulated using word-boundary probabilities estimated from child language corpora. Similar results were obtained for three different corpora. In our simulations, we also investigate the question whether children acquire their knowledge of word boundaries from words from the input, from the intake, or from their own output.

Introduction

The syllable is an important construct in phonological descriptions of languages (Van der Hulst & Ritter, 1999) as well as in models of language acquisition (Jusczyk, 1997) and language processing (Levelt, 1989). In most contemporary phonological theories the syllable plays an important role at the segmental level (e.g., in consonant harmony) as well as at the supra-segmental level (e.g., in stress assignment). Across languages syllables adhere to a number of universal principles (Venneman, 1988) and Clements (1990) proposes a universally valid algorithm for syllabifying words. One of its operating principles is 'sonority sequencing': a syllable has rising sonority from the left edge to the vocalic nucleus and falling sonority from the vowel to the right edge. Irrespective of the theoretical framework in which the universals of syllabification are cast, it is accepted that the language universals, such as those incorporated in Clements' algorithm, can be overruled by language-specific constraints. For instance, at the end of a syllable long vowels are universally accepted, but languages differ as to whether there can be a short vowel at the end of a syllable (Kager, 1989).

In sharp contrast to the relatively clear phonological picture stands the phonetic reality: what are the acoustic correlates of the syllable in the speech stream? For instance, acoustic correlates of the 'sonority sequencing principle' are very difficult to determine, which led phoneticians to define the syllable from a phonetic point of view as that entity of which the word *syllable* has three. The syllabic nucleus (the vowel) is fairly easy to detect, but the syllable boundaries are not straightforward. For instance, the /I/ in *bitter* is the nucleus of the first syllable, but where is the boundary of that syllable: immediately after the vowel /bI.t@r/ or after the first consonant /bIt.@r/ or in the middle of the first consonant /bIt.t@r/, a case of ambisyllabicity? This brings us to the core issue addressed in the present paper: if from a structural perspective syllables are easy to describe, but if it is very difficult to depict the acoustic correlates of the syllable and its boundaries, it is an outstanding question how children arrive at detecting syllables and their boundaries.

Nevertheless, in early speech perception (Jusczyk, 1997) as well as in speech production (Wijnen, 1988) children appear to use syllables as organizing entities. The question is: how does a child acquire the knowledge of the structure of syllables?

In the acquisition literature there are basically two approaches: in a nativist approach, the universals of syllable structure are thought to be innately given: they are described as inborn parameters (Fikkert, 1998), or as inborn constraints (Kager, 1999; Levelt, Schiller, & Levelt, 2000). Acquiring the structure of syllables requires a child to figure out the language-specific parameter setting or the language-specific constraint ranking. Thus, the broad outlines are genetically given, so that only on the basis of the ambient language the child has to determine where precisely her mother tongue fits into these outlines. Appealing as this may sound, it is unclear on what basis parameters are set or constraints are ranked. The cues for parameter setting or constraint ranking can only be found in the input. However, the acoustic correlates of the syllable are not clear in the input (see second paragraph).

The alternative approach is that children do not start from a preset body of knowledge, but instead use the information available in the input to arrive at linguistically relevant knowledge. For instance, Brent and Cartwright (1996) found that word boundaries can be learned on the basis of utterance boundaries. In a similar vein we want to investigate if syllable boundaries can be learned on the basis of word boundaries. Word boundaries are clear and usable cues, because words often occur in isolation in child-directed speech (van de Weijer, 1998; Brent & Siskind, 2001). Thus, the hypothesis tested in this paper is that syllable boundaries are learned on the basis of word boundaries.

In this paper we will test this hypothesis in a simulation experiment. The results of the simulations will be evaluated in the light of children's actual syllabification behavior. Gillis and De Schutter (1996) tested 5- and 6-year-old native Dutch-speaking children in a syllabification task: they syllabified disyllabic Dutch monomorphemes with a single intervocalic consonant, such as /Ap@l/ 'apple'. The children segmented the test words orally and of the possible syllabifications (V.CV, e.g. /A.p@l/, VC.V, e.g. /Ap.@l/, and VC1.C1V, e.g. /Ap.p@l/) the preferred syllabification pattern was V.CV (81.6%), i.e. before the intervocalic consonant. The next most frequent syllabification was the ambisyllabic pattern VC1.C1V (17.8%), and the children almost never (0.4%) put a syllable boundary after the intervocalic consonant (VC.V). Furthermore, children's syllabification of the intervocalic appeared to depend on the length of the preceding vowel, the stress pattern of the word, and the quality of the intervocalic consonant. These results will be taken as the background against which the results of the simulations will be evaluated.

Naive Bayesian learning of syllabification

Whether it is possible to learn syllable boundaries from information about word boundaries will be investigated with a naive Bayesian learning technique. A simple statistical model uses estimated word boundary probabilities of segments to predict syllable boundaries. This model takes into account the probability that a phoneme occurs at the end of a word and that the following phoneme occurs at the start of a word, and combines both features in a multiplicative way. Such a model does not take into account interactions between the features, hence Naive Bayesian learning, a well-known supervised learning approach (Mitchell, 1997). However, we don't use a normal supervised learning set-up in which training and testing is on the same data. In our case, training is on word boundary information and extrapolation is to syllable boundary decisions.

In the training data, the word-initial boundary probability and the word-final boundary probability of every phoneme are computed. This is done by counting the number of times the phoneme is at the end of a word (e), the number of times the phoneme is at the beginning of a word (b), and the total number of times that phoneme occurs (t). For every phoneme the word-initial boundary probability and the word-final boundary probability are then computed in the following way:

$$p(\text{beginning}(\text{phoneme})) = b/t$$
$$p(\text{end}(\text{phoneme})) = e/t$$

The model's task is to predict the syllable boundary in disyllabic monomorphemic Dutch words with one intervocalic consonant. To compute the probability of a syllable boundary between two phonemes of a test word, the word-final boundary probability of the first phoneme, and the word-initial boundary probability of the second phoneme are multiplied. This is done for the two possible syllabifications of the test word, considering that every syllable must contain a vowel.

E.g. *appel*, /Ap@l/, 'apple'

p(end(A))*p(beginning(p)) = p(V.CV)
p(end(p))*p(beginning(@)) = p(VC.V)

For all the test words the probabilities of V.CV and of VC.V are computed. I.e. the probability that the syllable boundary falls either before or after the intervocalic consonant. Ambisyllabicity occurs if the difference between those two numbers does not exceed a maximum limit. If it does, the pattern with the highest probability is chosen. This method forces the model to syllabify and to choose one of three syllabification patterns. No syllabification occurs, though, if the probability for both V.CV and VC.V is zero. This way, a fourth category of "no syllabification" is created, to make sure these cases are not counted as ambisyllabicity.

```
For $n$ = threshold:
    if        p(V.CV) = 0 and p(VC.V) = 0
              → no syllabification
    else if   |p(V.CV) - p(VC.V)| < n
              → VC1.C1V (ambisyllabic)
    else      max(p(V.CV), p(VC.V))
```

As the probabilistic model is trained on a two-way classification problem (either there is a word boundary or not), and the target classification problem is four-way (ambisyllabic, before or after the intervocalic consonant, no syllabification), we fixed the model on the proportion of ambisyllabicity found in the empirical data by setting the n threshold. This threshold value is determined by the amount of ambisyllabicity. The percentage of ambisyllabic syllabification is put as close as possible to 17.8%, which is the percentage of ambisyllabicity found in the experiments by Gillis and De Schutter (1996).

The fixing of a threshold parameter on the test data to be explained is an unfortunate consequence of the fact that the training data (word segmentation information) does not contain a similar concept to ambisyllabicity at the syllable level. Nevertheless, the threshold value seems to be rather robust over different training data sets, and could be learned with simple hill-climbing type of algorithms (there is a smooth gradient).

Research questions

Considering the different factors that might play a role in syllabification, a number of research questions were formulated.

1. What is the nature of the child's primary linguistic data? To acquire knowledge of language, children may analyze all the language that they hear or that is addressed to them (i.e., child-directed speech). Alternatively, it may well be that it is not the *input*, but the *intake* (i.e., what the child picks

   up from the input) (Wijnen, 2000) that is crucial for analysis. Alternatively, proponents of the output-as-input hypothesis (Elbers, 2000) argue that the input for children's linguistic analysis is primarily their own production, their own *output*.

2. What type of words is children's language analysis based on? Judging from the absence of function words in children's early productive vocabulary, it may well be that only content words are vital. And since syllables play a role in children's earliest word productions (Fikkert, 1998), it is important to investigate if syllabification can be acquired solely on the basis of content words as opposed to function words.

   Judging from the predominance of monosyllabic words in children's early production (or even the fact that all children initially exclusively produce monosyllables (Fikkert, 1998)) also the opposition between monosyllables and polysyllables will be investigated.

3. What is the influence of frequency on the acquisition of syllabification? Frequent words in the input are more salient for children (Jusczyk, 1997). However, Schreuder and Baayen (1997) found that the word frequency effect is composite in nature in the sense that it has both a token and a type component.

4. What is the optimal representation? Are words best represented as phonemes, or as phoneme categories? And is stress part of the representation?

   Phoneme categories express distinctive articulatory and acoustic features of phonemes, which is the reason why they differ in their scale of sonority. Sonority is regarded as important in syllabification, e.g. the universal Sonority Sequencing Principle describes syllables in terms of rising and falling sonority (Selkirk, 1984; Clements, 1990).

   Stress as well has been suggested as a determining factor in syllabification. There is a significant interaction between stress and length of the first vowel (Gillis & De Schutter, 1996), and there is less syllabification after the vowel if the first syllable is stressed than if it is unstressed (Wijnen, 1988).

In the following sections, we will report on experiments in which these dimensions are systematically encoded in the training data. The degree to which the resulting syllabification behavior of our statistical model matches the empirical data may have heuristic value to answer the question which dimensions of language data and representation are relevant in explaining this aspect of language acquisition.

Experiments

The input for the learner consisted of data taken from three Dutch child language corpora, all available through CHILDES (MacWhinney, 2000). The research questions were translated into different selections of input material and different types of input representations that were systematically varied in order to figure out their influence on the learnability of the task. Experiments were performed

1. using as training material the input to the child, the child's intake, and the child's output (the concept of intake was operationalized by using the actual adult model form of a child production, which makes intake a subset of the input);

2. using as training material different types of words: all words vs. content words, monosyllabic vs. polysyllabic words;

3. with information about word frequencies: word types vs. word tokens, as calculated from the corpora;

4. in which the representation of the input was varied: raw segmental material (phonemes) vs. segment categories (stops, fricatives, nasals, liquids, glides, and vowels) both with and without primary stress marking.

Combining all these factors in three child language corpora leads to a total of 136 experiments. In each case, the test material consisted of the words that were used in the experiment with children (Gillis & De Schutter, 1996) (see introduction). The artificial learner is set to the same task as the children: predicting the syllable boundary in Dutch disyllabic monomorphemes with a single intervocalic consonant. Hence, the learner has to decide whether for a given word (e.g. *appel*, /Ap@l/, 'apple') the string VCV should be syllabified as V.CV (/A.p@l/), VC1.C1V (/Ap.p@l/) or VC.V (/Ap.@l/).

For the different datasets, word boundary probabilities are computed with a naive Bayesian learning technique as described above. The amount of ambisyllabicity will be more or less the same for all the experiments (as close as possible to 17.8%), because the threshold (n), which is needed to get this percentage of ambisyllabicity, is dataset-specific. It is the percentages of syllabification after the vowel and after the intervocalic consonant, and the amount of "no syllabification", which are of interest. The results will be evaluated by comparing the proportions of the chosen syllabification patterns using word boundary probabilities to those of the children in the experiment by Gillis and De Schutter (1996). This means very little syllabification after the intervocalic consonant (0.4%) and most syllabification after the vowel (81.6%) are best.

Results

In this paragraph we will systematically take up the research questions formulated above and discuss what answer is suggested by the results of the simulation experiments. We will then propose the characteristics of the 'optimal' simulation, i.e., the one that most closely matches the results of the experiment with children.

Overall effects

1. What is the nature of the primary linguistic data?

   It is not clear from the simulation experiments' results whether language input, intake or production is the source of linguistic knowledge.

   Overall, there is less syllabification after the intervocalic consonant and less after the vowel in experiments using input or intake than in experiments using language output (Table 1).

Table 1: Average results over all simulation experiments using input vs. intake vs. output.

	V.CV	VC.V
input	51.9%	18.3%
intake	50.8%	16.4%
output	57.9%	23%

2. What type of words is the language analysis based on?

   The results suggest that content words — both mono- and polysyllabic — are the words used in a syllabification task.

   On average, there is less syllabification after the intervocalic consonant and more after the vowel in experiments using content words than using all words. The results of experiments using both mono- and polysyllabic words are better than those using only monosyllabic words. There is less syllabification after the consonant and more after the vowel with monosyllabic content words (types) than with all monosyllables (types or tokens), but there is more syllabification after the consonant and less after the vowel with monosyllabic content words (tokens) than with all monosyllables (Table 2).

Table 2: Average results over all simulation experiments using all words vs. content words vs. monosyllables vs. monosyllabic content words.

	V.CV	VC.V
content words	60%	6%
monosyll. content words types	51.1%	7.7%
all words	59.8%	19.8%
monosyllables	44.6%	30.1%
monosyll. content words tokens	17.3%	39.2%

3. What is the influence of frequency on the acquisition of syllabification?

   The simulation experiments suggest that linguistic analysis is based on word types rather than on word tokens.

   If information of word tokens is taken from child language corpora as training material, syllabification occurs more often after the intervocalic consonant and less after the vowel than when word types are used (Table 3).

Table 3: Average results over all simulation experiments using word types vs. word tokens.

	V.CV	VC.V
types	55.9%	14.2%
tokens	48.2%	21.5%

4. What is the optimal representation?

   A representation in phoneme categories appears to be more appropriate than a representation in phonemes.

   Using phoneme categories instead of phonemes generally gives better results, because with phonemes "no syllabification" is often assigned. The amount of test words for which the probabilities for V.CV and for VC.V are both zero can reach up to 81.1%. With phoneme categories, on the contrary, there are no test words that do not get syllabified (Table 4).

   The effect of stress marking in polysyllabic words is not univocal (Table 5). Using phoneme categories, there is less syllabification after the vowel and less after the consonant with stress marking; using phonemes, stress marking has the opposite effect. Thus, stress has a differential effect depending on the representation of the segments.

Table 4: Average results over all simulation experiments using a representation in phonemes vs. phoneme categories.

	V.CV	VC.V	no syll.
phoneme categories	68%	12.9%	0%
phonemes	36.1%	22.8%	22.5%

Table 5: Average results over the simulation experiments with polysyllabic words using a representation with vs. without stress marking.

	V.CV	VC.V
phoneme categories without stress marking	79.2%	4.3%
phoneme categories with stress marking	76.1%	4.1%
phonemes without stress marking	37.7%	21.6%
phonemes with stress marking	39.5%	22.2%

These tendencies concerning the composition and the representation of the input material are found over the total of all 136 experiments. Now we will discuss the individual experiments that most closely match the behavior of children.

Best results

Similar syllabification patterns (χ^2=1.16, p>0.05) to children's intuitive syllabifications in the experiments by Gillis and De Schutter (1996) are obtained when word boundary probabilities are computed in content words from the intake (types or tokens) or from the input (types) of a child language corpus, represented in terms of segment categories without stress assigned. These results are robust over the three language corpora, in the sense that we find the same results as displayed in figure 1 for the three corpora.

Not only the proportions of the syllabification patterns of the Naive Bayesian learner are similar to children's. Also the factors that influenced the children's syllabification patterns were replicated in the simulations. We will restrict the discussion to the factor of consonant quality.

Gillis and De Schutter (1996) found that children give significantly less ambisyllabic reponses if the intervocalic consonant is a stop (3.4%) than if it is a continuant (19.5%). Looking at the predicted syllable boundaries in the experiments, there are significantly less ambisyllabic responses as well if the intervocalic consonant is a stop (0%) than if it is a continuant (19.4%) (χ^2=9.05, p<0.01). This concerns the same training material, i.e. content words in phoneme categories from intake or input.

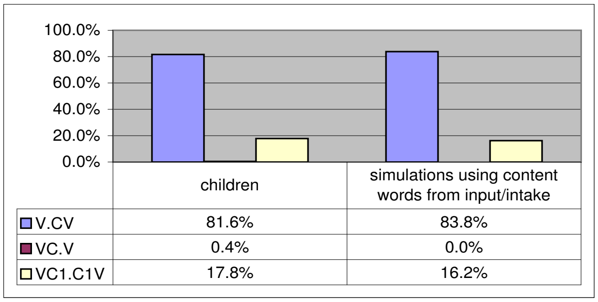

Figure 1: Comparison of syllabification patterns in five- and six-year-olds with results of simulation experiments using content words from input (types) or intake (types or tokens) in phoneme categories.

The observation that a simple statistical model trained on word boundary information (of content words in the input and using a representation in terms of phoneme categories) produces a tight fit with syllabification behavior in children, and the additional evidence that the model matches the children's behavior even at a detailed level of consonant quality is a strong existence proof of the possibility of data-oriented acquisition of the concept of syllables and of syllabification behavior.

Conclusion

Five- to six-year-old children that can't read nor write yet are able to syllabify disyllabic monomorphemic words according to universal rules of syllabification (Gillis & De Schutter, 1996). In this paper, we hypothesized that this intuitive knowledge of syllable boundaries is learned by attending to word boundaries.

To test this hypothesis, statistical word boundary probabilities of phoneme categories were used to predict syllable boundaries in disyllabic monomorphemes with one intervocalic consonant. To compute the probability of a syllable boundary between two phoneme categories, the word-final boundary probability of the first phoneme category and the word-initial boundary probability of the following phoneme category were multiplied. If the difference between the probabilities of the two syllabification possibilities (V.CV and VC.V) does not exceed a maximum limit, ambisyllabicity was assigned (VC1.C1V). Otherwise, the syllable boundary with the highest probability was chosen.

Using this naive Bayesian learning technique, similar syllabification patterns to children's intuitive syllabifications in the experiment by Gillis and De Schutter (1996) were obtained. Best results were achieved when the words used as material to compute word boundary probabilities were content words from the intake (types or tokens) or from the input (types) of a child language corpus, represented in phoneme categories. The same results were found with words from three different child language corpora. Moreover, the quality of the intervocalic consonant has a similar effect on children's intuitive syllabification and on the simulations using word boundary probabilities for syllabification. In both cases there is significantly less ambisyllabicity if the intervocalic consonant is a stop than if it is a continuant.

We have given an existence proof of the hypothesis that syllable boundaries can be learned from word boundaries. The fact that extrapolation from word boundaries to syllable boundaries can be modeled with such a simple statistical mechanism lends support to our initial hypothesis. Furthermore, varying the representations and input data used by this simple statistical learner, we were able to derive a number of interesting more detailed hypotheses about the type of representations and input children may use. More in particular, our results suggest syllable boundaries are most reliably learned from **content words**' boundaries. The semantic saliency of content words seems to be reflected in language production. Moreover, the best results are obtained using **phoneme categories**, rather than the phonemes themselves. This points at the role of sonority in the production of syllables. Phonological saliency is also shown to be an influencing factor, since disyllabic words with intervocalic stops are syllabified significantly differently from disyllables with intervocalic continuants. Finally, we found that the material that worked best to compute boundary probabilities are words from the **intake** or from the **input** of child language corpora. This suggests that children's productions — in this case intuitive syllabifications — could be based on their language input rather than on analysis of their own output. All these findings and predictions from the model have to be further investigated.

References

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93–126.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*, B33–B44.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Between the grammar and the physics of speech*. New York: Cambridge University Press.

Elbers, L. (2000). An output-as-input hypothesis in language acquisition. In P. Broeder & J. Murre (Eds.), *Models of language acquisition: inductive and deductive approaches*. Oxford: Oxford University Press.

Fikkert, P. (1998). The acquisition of Dutch phonology. In S. Gillis & A. De Houwer (Eds.), *The acquisition of Dutch*. Amsterdam: John Benjamins.

Gillis, S. & De Schutter, G. (1996). Intuitive syllabification: universals and language specific constraints. *Journal of Child Language*, *23*, 487–514.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kager, R. (1989). *A metrical theory of stress and destressing in English and Dutch*. Dordrecht: Foris.

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

Levelt, C. C., Schiller, N. O., & Levelt, W. J. (2000). The acquisition of syllable types. *Language Acquisition*, *8*, 237–264.

Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mitchell, T. M. (1997). *Machine learning*. Singapore: McGraw-Hill Companies.

Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*, *37*, 118–139.

Selkirk, E. (1984). On the major class features and syllable theory. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure*. Cambridge, MA: MIT Press.

Van der Hulst, H., & Ritter, N. A. (1999). Theories of the syllable. In H. Van der Hulst & N. A. Ritter (Eds.), *The syllable: views and facts*. Berlin: Mouton de Gruyter.

Venneman, T. (1988). *Preference laws for syllable structure and the explanation of sound change*. Berlin: Mouton de Gruyter.

Weijer, J. van de (1998). *Language input for word discovery*. Doctoral dissertation, Max Planck Institute for Psycholinguistics, Nijmegen.

Wijnen, F. (1988). Spontaneous word fragmentations in children: evidence for the syllable as a unit in speech production. *Journal of Phonetics*, *16*, 187–202.

Wijnen, F. (2000). Input, intake and sequence in syntactic development. In M. Beers, B. van de Bogaerde, G. W. Bol, J. de Jong, & C. Rooijmans (Eds.) *From sound to sentence – Studies on first language acquisition*. Groningen: Centre for Language and Cognition.