Human and machine syllabification in French: A comparison

Jeremy Goslin[1,2], Alain Content[3], Jean-Philippe Goldman[1], Uli H. Frauenfelder[1]

[1] Laboratory of Experimental Psycholinguistics, Université de Genève, Switzerland.
[2] Department of Computer Science, University of Sheffield, UK.
[3] Laboratory of Experimental Psychology, Université Libre de Bruxelles, Belgium.

ABSTRACT

To ascertain the merits of different phonetic syllabification algorithms, their performance was compared and contrasted both against each other, using lexical analysis, and against human syllable boundary placement, using first or second syllable repetition of a bi-syllable non-word. Perception results show that second syllable repetition showed far greater consistency than that of the first suggesting that the former condition is a more accurate measure of boundary placement. Comparison of human and algorithm syllable boundary placement showed high categorial accuracy for the Dell and Laporte algorithms whilst suggesting the use of multiple concurrent algorithms to produce a measure of confidence for each syllable boundary judgement.

1. INTRODUCTION

The problem of splitting words into their component syllables is one of importance to a number of fields, including psycholinguistics, linguistics, and artificial speech recognition. Whilst there are a number of algorithms for performing this function no reliable benchmark exists for their comparison either against each other, or with human syllable segmentation. In order to achieve such a measure five commonly used phonetic syllabification algorithms will be compared against each other, and against human participants. A description of those algorithms used can be seen on Table 1.

Algorithm	Description	Example Syllabification	
		/admår/	/åkstaz/
Dell (Dell, 1995)	*allows no more than one consonant in the coda of non-final syllables but considers only part of the OBstruent LIquid clusters (OBLI) class as indivisible onset clusters.*	/ad.mår/	/åk.staz/
Laporte (Laporte, 1993)	*divides consonant clusters just before last consonant in clusters, disallowing segmentation of OBLI clusters*	/ad.mår/	/åks.taz/
Coursil (Coursil, 1992)	*based upon Sonority Cycle (Clements, 1990).*	/a.dmår/	/åk.staz/
Peereman (Peereman, 1998)	*uses the opposite of sonority, consonanticity, uses different phoneme classification and gives special status to /s/.*	/a.dmår/	/åks.taz/
MOP (Kahn, 1976)	*Maximum Onset Principal, allows longest possible consonant onset provided it exists at the beginning of a word.*	/ad.mår/	/åk.staz/

Table 1: Description for target syllabification algorithms and examples of syllabification for 'admettre' and 'extase'.

2. LEXICAL ANALYSIS

In order to achieve a measure of similarity between our target algorithms they were applied to the phonetic transcriptions of all words found in the BDLEX French lexicon (~ 23000 words). This allowed an analysis of the syllabification strategies used by the various algorithms on commonly found intervocalic consonant singletons and clusters as a factor of the number of types and the number of tokens found for each type. This analysis found 432 different consonant singleton/cluster types in BDLEX with a total of 38549 occurrences (tokens) making an average of 1.68 singletons/clusters per word.

	Laporte	Peereman	Coursil	MOP
Dell	98.5 , 78.0	94.8 , 77.3	97.3 , 72.9	95.8 , 78.7
Laporte		93.6 , 74.5	97.0 , 80.6	94.6 , 80.0
Peereman			95.2 , 78.0	96.8 , 80.6
Coursil				95.6 , 74.8

Table 2: Percentage agreement between syllabification algorithms on intervocalic consonant singletons and clusters found in BDLEX (% of 432 types , % of 38549 tokens).

Table 2 shows that the levels of agreement are relatively high and well distributed; no two algorithms have significantly greater similarity than any others. Another point shown by this analysis is the difference in agreement between analysis by type and token. Agreement by token (on average 95.92%) is greater than that by type (on average 77.54%) showing that algorithms agree more often on common clusters/singletons.

Concurring Algorithms	By Type	By Token
2	100.00	100.00
3	97.68	99.97
4	76.85	95.14
5	56.71	92.24

Table 3: Percentage of cluster/singleton types & tokens covered by different levels of algorithm agreement.

This effect can also be seen on Table 3. It is also interesting to note that over 50% of all clusters/singletons types are syllabified similarly by all algorithms (over 92% of tokens). This high level of agreement suggests a possible strategy in the use of these algorithms is not to use them individually, but concurrently. Thus the levels of agreement between algorithms can be used to impart a measure of *confidence* in the syllable boundary placement.

3. PERCEPTION EXPERIMENT

The purpose of the perception experiment is twofold, to form a benchmark measurement for the target syllabification algorithms, and to discover the consistency of the participants' syllable boundary placement over a wide range of possible intervocalic consonant clusters/singletons. To avoid lexical influences (Treiman & Danis, 1988) only non-word stimuli were used in this experiment.

3.1 Method

3.1.1 Stimuli

Using a subset of phonotactically legal intervocalic consonant singleton/cluster types from the lexical analysis, a list of stimuli were generated with between 1 and 3 intervocalic consonants. In order to create a bi-syllabic non-word, a random selection of vowels was placed at the start and end of each token (forming VCV, VCCV, and VCCCV stimuli) taken from either 'u', 'i', 'a', or 'y' (vowels thought not to affect syllable segmentation). Stimuli were organised using robust feature classes Nasal (N), Fricative (F), Liquid (L), and Plosive (P) into four singleton, 16 double (e.g. FN), and 3 triple consonant categories (e.g. PFP). Those categories contained between 1 and 3 different clusters with 6 tokens per category. This organisation of stimuli resulted in 138 tokens, with 58 clusters/singletons, and 23 feature categories, with an additional 20 tokens for training.

3.1.2 Procedure

Participants were asked to repeat either the first or second part of the bi-syllabic stimuli as quickly as possible. Experimental stimuli were arranged into 3 blocks, each of which was presented under two conditions, that is, the repetition of either the first or second syllable of the stimuli. Stimuli order for each block was randomised for each presentation. Stimuli in each block were presented on a continuous basis, one every two seconds without pause until the end of the block, where participants were invited to take a short pause. The experimental condition alternated on each successive block, with blocks ordered such that block repetitions (for the 1st and 2nd condition) were never adjacent to each other. At the start of the experiment, two short training blocks of 10 stimuli were presented, one for each experimental condition.

3.1.3 Participants

All 22 participants were students of the Université de Genève and were native speakers of French with no known hearing defects. They received course credits for their participation.

3.2 Results

One of the clearest results from the experiment was the difference in syllable boundary placement consistency between the onset and offset conditions. As can be seen in Figure 1, the percentage of participant responses for the preferred boundary placement, that is the segmentation given by the majority of participants, are far higher for condition 2 than condition 1 for the majority of feature categories (average consistencies of 83.6% and 92.4% for conditions 1 & 2 respectively), with all results from the former condition above levels of statistical significance (using Chi^2 measure) at $p < 0.05$. This suggests that the repetition of the second syllable of the stimuli maybe the more reliable measure of human syllabification. This assumption is also supported by findings of syllable boundary placement using single intervocalic consonant stimuli. These results are of particular interest because diverse models of syllabification, including the ‘obligatory onset principle’ (Hooper, 1972), are unanimous in their placement of the syllable boundary in the /V.CV/ position. The fact that over 99% of subjects agree with this placement for the second condition, whilst up to 40% disagree for the first also suggest that results of the second condition give a more accurate measure of syllabification.

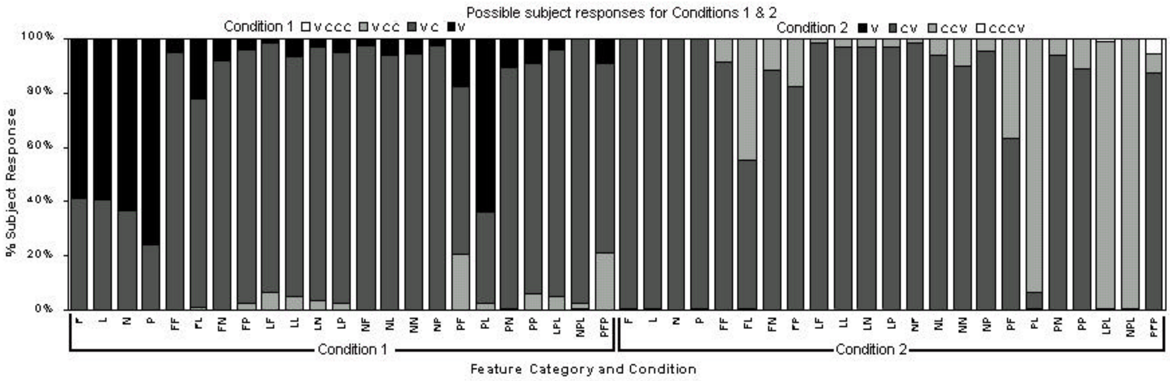

Figure 1: Graph of human syllabification responses for experimental conditions 1 & 2 for each feature category.

Similar differences in consistency between the repetition of the first and second parts of stimuli were also found in an unspeeded study of syllable boundary placement (Content et al, 1998). In this study such effects are cited as evidence against the boundary view of syllable segmentation. Instead, an ‘onset hypothesis’ is suggested in which distinct operations are used to detect syllable onsets offsets, with the former providing reference points for segmentation and lexical access.

Additional parallels may also be drawn from a study on syllable segmentation in pre-literate children (Floccia et al, 1999). This showed that the preferred syllable boundary placement in children are identical to our own, however, it is noted that consistency levels for our adult subjects were significantly higher than those of the syllable repetition task in children.

4. COMPARISON OF HUMAN AND AUTOMATIC SYLLABIFICATION

A comparison of human and automatic syllabification was made by applying each of our target algorithms to the stimuli used in the human perception experiment and comparing their responses to that of the participants responses under condition 2, that which has been shown to be the more reliable measure of human syllabification. Table 4 shows a summary of these results as a percentage of agreement between the algorithm and the participant’s majority decision for each type (each decision found to be statistically significant at $p < 0.05$). As can be seen it is the Dell and Laporte algorithms which perform most similarly to the responses of the perception experiment, agreeing with the vast majority of stimuli (Dell disagreeing on the PFP class, Laporte with that of the /vr/ cluster in the FL category). It is interesting to note that it is the two algorithms based on French specific phonotactic regularities, Dell and Laporte, which significantly outperform language generic methods.

Syllabification Algorithm				
Dell	Laporte	Peereman	Coursil	MOP
95.7	97.9	84.8	80.3	78.3

Table 4: Agreement levels (%) between syllabification algorithms and perception experiment results.

Confidence Level		
3	4	5
81.92	86.72	97.12

Table 5: Average consistency (% of participant responses) for projected confidence levels.

Comparing the participants syllabification consistency for each type with their projected confidence values, calculated during lexical analysis, we can see from Table 5 that there is a positive correlation between participants consistency and the degree of agreement between algorithms. Indeed, if this data is extrapolated to the whole of the consonant cluster/singleton set it seems that over 92% of token occurrences can be predicted with over 96% participant consistency. This information is useful in predicting which tokens, such as PL, FP, FL, and PF categories in our experiment, present the greatest problems in consistent syllable boundary placement, and as such merit further study.

5. CONCLUSION

From our experimental results it appears that the majority response of subjects' syllable boundary placement may be relatively accurately predicted using two (Dell and Laporte) of our tested algorithms. However, large differences in syllable boundary consistency values between our experimental conditions, coupled with the relatively high levels of consistency for some of our categories, suggest that phonetic rule based syllable boundary placement may not hold the key to human syllabification, and that further research into possible reasons for these inconsistencies will be required for a more accurate model of syllabification.

5. REFERENCES

Clements, G. N. (1990). The role of sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), Between the Grammar and Physics of Speech. New York: Cambridge University Press, pp. 283-333.

Content, A., Kearns, R. K & Frauenfelder, U. H. (1998). Boundaries versus onsets in syllable segmentation. Submitted for publication.

Coursil, J. (1992). Grammaire Analytique du Français Contemporain. Unpub. Diss., University of Caen.

Dell, F. (1995). Consonant clusters and phonological syllables in French. Lingua, 95, p.5-26.

Floccia, C., Goslin, J., Bouketir, N., Bradmetz, J. (1999). Consonant cluster syllabification in preliterate children. Proceedings of the 2nd Journées d'Etudes Linguistiques, Nantes.

Hooper, J. B. (1972). The syllable in phonological theory. Language, 48, p. 525-540.

Kahn, D. (1976). Syllable-based Generalisation in English Phonology. Unpub. Diss, MIT.

Laporte, E. (1993). Phonetic syllables in French: combinations, structure, and formal definitions. Acta Linguistica Hungarica, 41, pp, 175-189.

Peereman, R. (1998). Personal communication.

Treiman, R. & Danis, C. (1988). Syllabification of intervocalic consonants. Journal of Memory and Language, 27, pp. 87-104.

ACKNOWLEDGEMENTS

This work was supported by the Swiss National Fund for Scientific Research (grant no. 1113-04969896).