Content-agnostic online segmentation as a core operation

Federico Adolfi (federico.adolfi@esi-frankfurt.de)
ESI Neuroscience, Max-Planck Society, Germany

Yue Sun (yue.sun@esi-frankfurt.de)
ESI Neuroscience, Max-Planck Society, Germany

David Poeppel (david.poeppel@nyu.edu)
Department of Psychology, New York University, USA

Abstract

We approach the problem of explaining segmentation — the human capacity to partition input streams into representations of appropriate form and content for efficient downstream processing — by exploring a theoretically minimalistic and computationally plausible account of phoneme-to-word chunking. Through computational models, mathematical proofs, algorithm design, and observer model simulations in two languages, we suggest that online segmentation can be guided by content-agnostic properties of internal memory structures (i.e., lexicality and length type frequency). Our theoretical and empirical findings point to a formal link between such properties with practical performance benefits. Together, these contributions make progress on a fully explicit computational- and algorithmic-level account with plausible implementational-level primitives.

Keywords: high-level cognition; segmentation; computational modeling; algorithmic modeling; plausibility constraints; computational complexity; theory; mathematical analysis; simulation.

Introduction

Segmentation — the capacity to partition inputs into chunks of appropriate form and content for efficient downstream processing (Adolfi, Wareham, & van Rooij, 2023) — is a core cognitive subcomputation invoked in many fields of the cognitive sciences (most notably, speech and language, and also music; Poeppel & Assaneo, 2020; Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Cutler, 1994). In contrast to efforts involving expressive computational architectures leveraging multiple sources of information (e.g., Perruchet & Vinter, 1998; Friston et al., 2021; Swingley & Algayres, 2024), here we approach the problem of explaining segmentation with a theoretically minimalistic research program focused on cognitive plausibility: how accurately can a cognitive system segment inputs using as little knowledge as possible? Answering this question would pave the way not only to explain fully developed systems but also how they could develop and learn.

With this minimalistic imperative in mind and using *word segmentation* as a case study, we explore the plausibility of *content-agnostic* features of segments (i.e., blind to pragmatics, semantics, syntax, etc.) as the sole objectives guiding online processing. Building on proofs of tractability of certain classes of objective functions for segmentation (Adolfi, Wareham, & van Rooij, 2023), we consider a number of such features combined in tractable ways. Among them, we pay special attention to *lexicality* (whether a segment is represented in the internal memory system), *parsimony* (the number of boundaries needed to segment optimally, given environmental statistics), and *length type frequency* (the typicality of a segment length in the lexicon).

Here we suggest that these superficially dissimilar features share underlying properties with ties to human-aligned segmentation objectives. Moreover, we show that the synergy of lexicality and length type frequency is empirically superior to other properties such as *word token frequency*, *length token frequency*, or combinations thereof, in guiding segmentation. We support this hypothesis with computational models, mathematical proofs, algorithm design, and observer model simulations in 2 languages with different typological origins.

Overview. We first explain **[§1] A Minimalistic Approach to Segmentation** to study the smallest set of content-agnostic properties that can successfully guide online computation. In **[§2] Computational-level Modeling**, we propose candidate computational models of segmentation with content-agnostic objective functions, each of which fully specifies the computation mathematically. We conduct **[§2.2] Mathematical analyses** and prove that these, but not all conceivable objectives, preserve the optimal substructure of the problem that allows for segmentation computations to be reused efficiently as new information comes in. We also derive a close formal relationship between lexicality, parsimony, and length probability as objective functions, such that optimizing one implicitly optimizes the other. Based on this, in **[§3] Algorithmic-level Modeling** we propose an *online* algorithm for segmentation that leverages this structure using Dynamic Programming. Via **[§3.2] Complexity analyses** we prove it matches the target computation: it is guaranteed to find the optimal solution to any input and its running time scales feasibly with input size, as required for any cognitively plausible account. We deploy the algorithm in **[§4] Observer Model Simulations**, generating artificial segmentation behavior (phonemes-to-words) and comparing it to natural human outputs via a language corpus to assess the empirical adequacy of our theoretical analyses. In **[§5] Results & Discussion** we situate and interpret our findings. We show empirically that (i) optimal segmentations match human segmentations reasonably well in terms of performance, (ii) performance is best when optimizing lexicality and length type frequency, and (iii) relative performance relates to formal links between objectives.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

Together, these contributions advance a cognitively plausible, minimalistic theory of segmentation that is fully specified at the computational and algorithmic levels of explanation, and whose primitives, which have close ties to reinforcement learning and graph neural networks, can be readily compared to brain dynamics for implementational-level explanations.

§1 A Minimalistic Approach to Segmentation

We approach the problem of explaining segmentation through a *minimalistic* perspective. That is, we formulate consistent computational and algorithmic theories and models, with a minimal set of assumptions and an architecture limited by cognitive and complexity-theoretic plausibility constraints (Adolfi, 2023). To that end, we test theoretically and empirically the use of content-agnostic properties of internal memory structures and input regularities (see Table 1) as the sole guides to segmentation. We deploy this framework, CONTENT-AGNOSTIC ONLINE SEGMENTATION **(CAOS)**, to explain phoneme-to-word segmentation (see Figure 1).

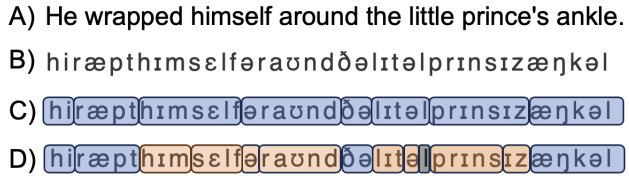

Figure 1: Phoneme-to-word chunking as a segmentation problem. (A) orthographic representation of (segmented) sentence. (B) unsegmented phoneme sequence (input to the segmentation computation). (C) correctly segmented sequence (output of an empirically optimal computation). (D) incorrectly segmented sequence illustrating possible biases inherent in objective functions for segmentation (output of a formally optimal but empirically suboptimal computation).

§2 Computational-level Modeling

We model the problem of segmentation as the process of constructing a chunking scheme for an input sequence such that the resulting segments are optimal according to cognitively plausible criteria. The following are some basic definitions we will use in the computational problem description.

§2.1 Computational problem

Definition 1 (Sequence)**.** We denote a finite sequence of elements as $S = (s_1, s_2, ..., s_N)$, and its length $|S| = N \in \mathbb{N}$, and basic units s_i.

Definition 2 (Segments and segmentation)**.** Given a sequence S, a segmentation of S is a sequence of contiguous subsequences called segments, $P = ((s_1, s_2, ...), ..., (..., s_{N-1}, s_N))$, where segments are disjoint, $\forall p_i, p_j \in P : p_i \cap p_j = \emptyset$, and span the original sequence, $\bigcup_{i=1}^{|P|} p_i = S$.

The problem of finding a segmentation P of S that maximizes some value $V(P_S)$,

$$\arg\max_{P_S} V(P_S), \quad P_S \in \{P_S^i\}_{i=1}^{2^{|S|}} \tag{1}$$

can be described as a computational problem, as follows.

Problem 1. SEGMENTATION (SEG)

Input: a sequence S, and a value function $V : P \mapsto V(P) \in \mathbb{R}$ that maps candidate segmentations to their respective quality.

Output: a segmentation of S such that its value $V(P)$ is maximum.

Note that not all possible objective functions $V(\cdot)$ will give rise to tractable or otherwise plausible computational-level models for segmentation in real-world cognitive systems, even prior to empirical testing (i.e., theoretically; Wareham, 1996). Consider, for instance, that the space of possible segmentations is exponential, and therefore if there is no internal structure in the mapping V that a system could exploit, there can be no efficient algorithmic-level counterparts to such a computational-level theory. On the other hand, making modest assumptions about the composition of the segmentation value yields tractable objectives (Adolfi, Wareham, & van Rooij, 2023). We build on these results to consider various such objective functions that a cognitive system could plausibly leverage, with a minimalistic view in mind.

Definition 3 (Segment scoring)**.** The value of a segmentation can be derived, in some cases, from the value of individual segments, namely, $V(P) = \sum_{p \in P} F(p)$, where $F : \mathscr{P} \to \mathbb{R}$ maps subsequences $p = (s_i, s_{i+1}, ..., s_{i+q})$ to a value $F(p)$.

Segment scoring functions can formalize various kinds of content-agnostic knowledge that the cognitive system might possess about segments (e.g., their length). We study knowledge of the environment statistics and those of internal memory structures (see Table 1). We write a segmentation problem with the set of objectives $\mathcal{V}$ as $\mathcal{V}$-SEG.

Definition 4 (Segmentation parsimony)**.** Given a segmentation P of sequence S, its parsimony is $M(P_S) = \frac{|P|}{|S|}$. A segmentation is more parsimonious if it partitions the sequence using fewer boundaries.

§2.2 Mathematical analyses

Here we derive formal relationships between *lexicality*, *parsimony*, and *length* as objectives. Given a sequence S, assume possible segments are indexed by i, and consider an objective function with segment values and a parsimony regularizer,

$$V(P_S) = \sum_{i=1}^{|P|} F(p_i) x_i - \alpha \sum_{i=1}^{|P|} \frac{x_i}{|S|} \tag{2}$$

where $x_i \in \{0, 1\}$ is a binary variable that indicates the inclusion of segment i in the segmentation, $F(p_i)$ is the value of segment i according to some scoring function F, and $\alpha \in \mathbb{R}_{\geq 0}$ is a hyperparameter for the relative importance of objectives.

The first summation contains terms dependent only on lexical segments (i.e., where $F(p_i) > 0$), and the second contains both lexical and non-lexical terms. With this, we can derive the following form that segregates lexical and non-lexical terms corresponding to the candidate segmentation.

$$V(P_S) = Q_l(1 - \frac{\alpha}{|S|}) - Q_{nl}\frac{\alpha}{|S|} \tag{3}$$

where Q_l and Q_{nl} are quantities proportional to the number of segments in the segmentation P_S of S that are lexical (l) or non-lexical (nl), respectively, and α sets their contribution.

From this form of the objective we can easily read out that a regularizer encouraging parsimony will penalize non-lexical items in proportion to α. This is most readily appreciated when letting $F(\cdot)$ be the lexicality, $F(p_i) \in \{0,1\}$, of the segment p_i. This establishes a formal relationship between *lexicality* and *parsimony*, as defined above. It also derives from an objective with mixed segment- and segmentation-specific values an objective purely as a function of segment-specific values. This is important as the latter shows that it preserves the optimal substructure of the optimization problem that will be exploited later by our algorithmic-level theory.

Segmentation parsimony naturally relates to the length of the chosen segments, as clearly

$$\frac{1}{|P|}\sum_{i=1}^{|P|}|p_i| = \frac{1}{M(P_S)}. \tag{4}$$

This will be important later on as we consider (a) the effect of *parsimony* as a regularizer on objectives such as *word token frequency*, where biases can complement each other, and (b) the effect of objectives such as *length type frequency*.

Content-agnostic property	Notation
System	
Lexicality	Lex
Length type frequency	L_{Freq}^{Type}
Environment	
Word token frequency	W_{Freq}^{Token}
Parsimony	$Pars$
Length token frequency	L_{Freq}^{Token}

Table 1: Content-agnostic properties[1].

§3 Algorithmic-level Modeling

Here we present an algorithm for online segmentation and prove that it computes the segmentation problem exactly and efficiently (i.e., it finds optimal solutions to any instance in a feasible amount of time).

§3.1 Algorithm design

We use the *optimal substructure* property of the computational problem to design a dynamic programming, online algorithm. Algorithm 1 describes the computation at any given time when an input unit is received. It generates a structured subset of candidate segmentations based on optimal solutions obtained in previous steps and a new sequence unit obtained in the current step. Only these, among exponentially many, are evaluated, and the segmentation which maximizes the objective is kept as a solution at the current step.

[1] *Token* and *Type* designate properties of the environment and internal memory, modeled by the corpus and lexicon, respectively.

Algorithm 1 — Construct optimal segmentation online.

Input:

s_N ▷ `input unit at time N`

$V : P \mapsto V(P) \in \mathbb{R}$ ▷ `value function`

$\mathcal{P}_{N-1} = \{(P_t^*, V_t^*)\}_{t=0}^{N-1}$ ▷ `previous solutions & values`

s.t. $P_t^* := ((s_1, \ldots), \cdots, (\ldots, s_t))$ ▷ `solution at time t`

Output:

$\mathcal{P}_N = \{(P_t^*, V_t^*)\}_{t=0}^{N}$ ▷ `Solutions & values up to N`

procedure ONLINESEGMENTATION(s_N, V, $\mathcal{P}_{N-1}$)
  P $\leftarrow$ P^*_{N-1}**.concat**$((s_N))$
  V $\leftarrow$ V(P)
  for $t \leftarrow 1$ to $N-1$ **do**
    candidate $\leftarrow$ P^*_{N-1-t}**.concat**$((s_{N-t}, \ldots, s_N))$
    value $\leftarrow$ V(candidate)
    if value $>$ V **then**
      P $\leftarrow$ candidate
      V $\leftarrow$ value
    end if
  end for
  $\mathcal{P}_N \leftarrow \mathcal{P}_{N-1}$**.append**$((\text{P}, \text{V}))$
  return $\mathcal{P}_N$
end procedure

§3.2 Complexity analyses

Computations that cannot be done efficiently and reliably for the relevant domain of inputs are implausible as cognitive models (Wareham, 1996). It is important to show that computational-level models admit efficient algorithms and that algorithmic-level models are both efficient and provably match the proposed computation. The following definitions formalize these basic complexity-theoretic plausibility constraints on computational and algorithmic cognitive models.

Definition 5 (Polynomial-time tractability)**.** An algorithm is said to run in *polynomial-time* if the number of steps it performs is $O(n^c)$, where n is a measure of the input size and c is some constant. A problem Π is said to be *tractable* if it has a *polynomial-time algorithm*. P denotes the class of such problems.

Conjecture 1. P $\neq$ NP.

Definition 6 (Polynomial-time intractability)**.** The class NP contains all problems in P and more. Assuming Conjecture 1, *NP-hard* problems lie outside P. These problems are considered *intractable* because they cannot be solved in polynomial-time (unless Conjecture 1 is false; see Fortnow, 2009).

We now formally prove that both our computational and algorithmic models adhere to these constraints, and the latter constitute exact solutions to the former.

Proposition 1. Algorithm 1 computes $\mathcal{V}$-SEG exactly, where $\mathcal{V} \subseteq \{L_{Freq}^{Type}, Lex, W_{Freq}^{Token}, Pars, L_{Freq}^{Token}\}$.

Proof (sketch). Consider the base case: the algorithm gets as input a sequence unit, and the previous optimal solution, which is by default the empty segmentation. In subsequent steps, the algorithm need only compare the value of segmentations that involve the new input unit as part of a new possible segment of size t and the complementary optimal solution of size $N-t$. Alternative solutions must be suboptimal. ■

Proposition 2. Algorithm 1 is polynomial-time.

Proof (sketch). Note that even though there are $2^{|S|}$ possible segmentations of a sequence S, Algorithm 1 exploits the structure of the objective to compare only $O(t)$ candidate segmentations at each time step t, evaluating $N(N+1)/2$ candidates overall when considering all input units $s_i \in S$. The whole procedure runs in $\sim O(N^2)$, where $N = |S|$. ■

Corollary 1. $\mathcal{V}$-SEG is efficiently computable (i.e., in PTIME), where $\mathcal{V} \subseteq \{L^{Type}_{Freq}, Lex, W^{Token}_{Freq}, Pars, L^{Token}_{Freq}\}$.

These analyses establish the a-priori, complexity-theoretic plausibility of our proposed computational- and algorithmic-level models, their mutual consistency, and the feasibility of online computation through dynamic programming, as explanations for segmentation in real-world cognitive systems.

§4 Observer Model Simulations

Although we proved various desirable properties of our computational- and algorithmic-level models, it is still possible to have a misalignment of formal objectives and human-like segmentation (Adolfi, van de Braak, & Woensdregt, 2024). To test the empirical adequacy of our theoretical models we need a process model capable of simulating segmentation behavior under the conditions analyzed formally. This allows us to compare artificial segmentation guided by the formalized objectives against human-like segmentation on natural language inputs. We implement an observer model out of the aforementioned components and a lexicon, and simulate its behavior on a language corpus under various combinations of objective functions[2]. To assess the robustness of our results and inferences, we run all simulations in two languages with different typological origins: *English* and *Chinese*.

§4.1 Corpus

The environment of the observer (i.e., the input domain to sample from) is modeled by a language corpus. Our set of inputs for the observer model consists of sentences from the English and Chinese text Little Prince curated in Le Petit Prince fMRI Corpus (Li et al., 2022). For each corpus, we phonemically transcribed every word using a lexical database of the same language (Table 2), such that each sentence could be converted into a sequence of phonemes.

[2]In the case of *parsimony* we find the best α via grid search.

Language	Corpus	Lexicon
English	1499 sentences (15604 words)	WebCelex 70849 wordforms
Chinese	1577 sentences (19587 words)	Chinese Lexical Database 48826 wordforms

Table 2: Corpus and lexicon materials.

§4.2 Lexicon

The internal knowledge of the observer is modeled by a lexicon (Table 2), the cognitive basis for computing values of segments and segmentations (Definition 3). We used the English Celex database (`http://celex.mpi.nl/`) and the Chinese Lexical Database (C. C. Sun, Hendrix, Ma, & Baayen, 2018). The frequency of occurrence of each word was taken from the SUBTLEX corpus for each language, which estimates word frequencies from movies and TV series subtitles (Brysbaert & New, 2009; Cai & Brysbaert, 2010).

§4.3 Word length frequency

To establish the distribution of word length across all words, we computed, for each language, the frequency of occurrence of each word length (i.e., number of phonemes) in the lexicon (type frequency) and in real-world sentence corpora (token frequency). For each lexicon, we computed the *type frequency*, (i.e., the number of distinct words that have a specific length). Type frequency does not take into account how often words with different lengths are used in real-world language communication. This frequency indicates the relative abundance of words of different lengths in the lexicon. We also computed the *token frequency* for each word length, which is based on the number of occurrence of words of each length in the SUBTLEX corpus for each language. This quantifies how often words of a given length appear in real-world sentences.

§4.4 Simulations

We run Algorithm 1, augmented with the lexicon described above, on the corpus and plot performance measured by WER for various combinations of objectives[3] in Figure 2 and 3.

Task. The observer model is presented with unsegmented sequences of phonemes sampled from the corpus and is tasked with segmenting them (see Figure 1) according to its internal objective function. Its performance, however, is evaluated against the ground truth segmentation in the corpus.

Performance evaluation. We measure the performance of observer models using *word error rate* (WER). It is derived from the number of substitutions S, deletions D, insertions I, and correct words C. $\text{WER} = \frac{S+D+I}{S+D+C}$. Lower scores indicate fewer errors and hence better performance[4].

[3]For conditions including parsimony (*Pars*), we plot the performance for the best nonzero weight of the parsimony term as a separate condition even when it is worse than without parsimony.

[4]Note that the numerator includes insertions, and $N_{ref} = S+D+C$ is the number of words in the reference sentence, which appears in the denominator. Therefore, WER can be greater than 1.

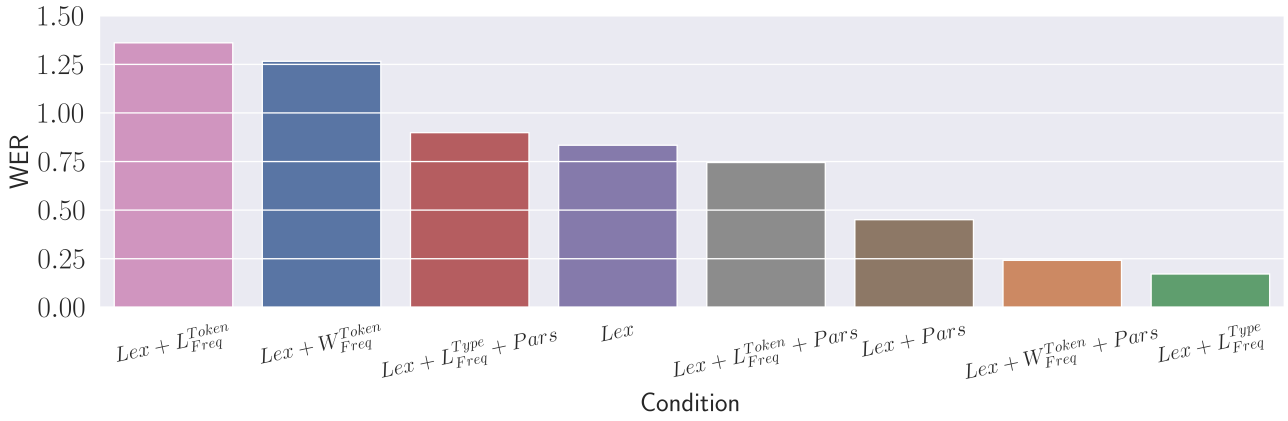

Figure 2: Results for English quantified as WER for each combination of objective. The combination of lexicality and length type frequency, features of the internal memory structure (lexicon), attains the best segmentation performance ($\sim$ 18%).

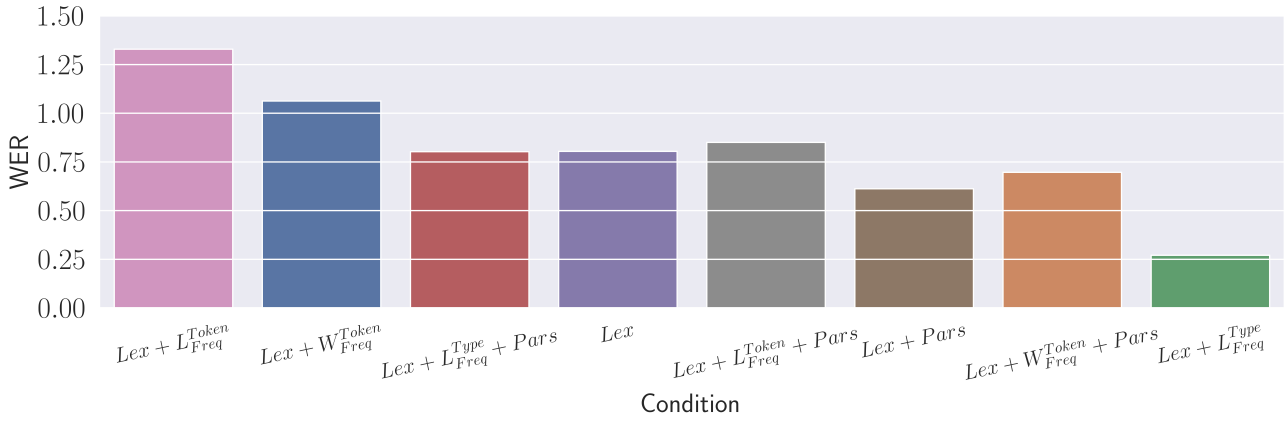

Figure 3: Results for Chinese quantified as WER for each combination of objective. The best performance ($\sim$ 25%) is achieved by the objective combining the same features of the internal memory structure (lexicon) as in English, namely, lexicality and length type frequency.

§5 Results & Discussion

Overall, our findings show that cognitively plausible, content-agnostic summary statistics are viable objective functions for segmentation, and shed light on the formal relationship between *lexicality*, *parsimony*, and *length type frequency* [§2.2]. In particular, the length type distribution is theoretically viable, as it preserves optimal substructure in the segmentation problem [§2.1] that allows for exact and efficient [§3.2] online algorithms based on dynamic programming [§3.1], and it is empirically superior in terms of segmentation performance across two languages. These results are robust across two languages of different typological origin (Enlgish, Figure 2; and Chinese, Figure 3). The following subsections further situate and interpret the results obtained.

The goal of segmentation. The primary goal of word segmentation is to parse a continuous sequence of elementary sound units (e.g., phonemes) into linguistically meaningful units (words). A key component of successful segmentation is to ensure that the identified segments correspond to entries in the lexicon — the objective of *lexicality*. An effective segmentation strategy produces segmented sequences that *only* contain lexical segments (i.e., present in an ideal lexicon).

The issue of over-segmentation. Results from the lexicality condition showed mediocre performance (*Lex*; Figure 2 and 3). A segmentation strategy guided by lexicality alone would inherently favor segmentation solutions that contain more lexical segments. Therefore, this strategy is intrinsically susceptible to the issue of over-segmentation (see Figure 1D for an example). In English, short lexical segments are often also part of longer words. Under a greedy segmentation objective that maximizes lexicality, long words would be (often erroneously) parsed into a series of short words, possibly interspersed with non-lexical segments. While such a segmented sequence would perhaps contain many lexical items, over-segmentation can lead to many non-lexical segments (labeled as 'insertions' by measures such as WER).

Word frequency as sole objective: a poor guide. An intuitive idea to improve segmentation performance is by incorporating environmental (i.e., input) statistics into objectives for segment recognition. A natural candidate is word fre-

quency, under the assumption that frequent words should be easier to recognize, since they appear more often in speech. This idea is well supported by empirical findings that frequent words are recognized faster and more accurately than rare words (Rubenstein, Garfield, & Millikan, 1970; Ferrand et al., 2018). We tested this in the word token frequency condition. However, our results showed that segmentation performance worsens when lexicality is weighted by word token frequency ($Lex + W_{Freq}^{Token}$). This outcome could be explained by the inverse relation between word frequency and word length (i.e., Zipf's law Zipf, 2016). Since shorter words tend to be used more frequently in natural language, a segmentation strategy that favors segments with higher frequencies can only exacerbate the over-segmentation problem. That is, when the system is biased toward selecting segmentations with short words, it further partitions long words into multiple short lexical and non-lexical segments.

Parsimony to counter over-segmentation. Given the issue of over-segmentation, a cognitively plausible hypothesis is the existence of a 'parsimony' regularizer — a counterforce towards segmentation solutions that minimize the number of segments. Parsimony has been widely applied in various fields beyond linguistics (e.g., Beekhuizen, Bod, & Zuidema, 2013), as a general heuristic favoring simpler, more efficient structures (Sober, 1981). Results for the observer model with such a regularization scheme show a substantial improvement in performance. The combination of lexicality and parsimony ($Lex + Pars$) yielded much better results than lexicality alone (an improvement of $\sim$ 20-35%). More strikingly, pairing word token frequency with parsimony not only corrects the over-segmentation problem but actually outperforms the lexicality-parsimony objective ($Lex + W_{Freq}^{Token} + Pars$). This suggests that parsimony fits well as a counterforce to the inherent biases in objectives such as word token frequency.

Segment length: type versus token frequency. We conjectured that parsimony improves segmentation by counteracting the over-segmentation of long words. On this view, its effectiveness derives from a regularization mechanism whose implicit basis is word length distribution. This is the motivation for studying an explicit word length regularizer. We explored two of kinds objectives that embody different perspectives on word length constraints: (i) *length type frequency* — the distribution of word lengths in the system's internal knowledge structure of the language, modeled by the lexicon; (ii) *length token frequency* — the frequency of different words lengths in real-world language use, modeled by corpus-based statistics. Interestingly, our results revealed that the best performance for each language ($\sim$ 18% error rate for English and $\sim$ 25% for Chinese), is achieved with the lexicality and length type frequency condition ($Lex + L_{Freq}^{Type}$), while the lexicality and length token frequency objective ($Lex + L_{Freq}^{Token}$) performs very poorly (Figure 2 and 3). This finding emphasizes that, although intuitively the best performance should come about from tuning the system to input statistics (the material the system is tested on), summary statistics of internal memory structures (e.g., length distribution in the lexicon) can provide the optimal balance of inductive biases. This is underexplored in the interpretability and alignment of natural and artificial segmentation (Adolfi, Bowers, & Poeppel, 2023).

Word token frequency versus length type frequency. An interesting observation is that word token frequency combined with parsimony achieved worse but similar performance as compared to length type frequency. One key difference, however, is that length type frequency requires only lexicon-based knowledge, whereas word token frequency and parsimony relies on environmental input statistics, introducing additional complexity. From a minimalist perspective, length type frequency may be preferable. This suggests that the cognitive system's sensitivity to word length is not dictated merely by the most frequently used words but rather by the typical word lengths that carry the most relevance to the internal organization of word forms and meaning in the lexicon (e.g., Dautriche, Mahowald, Gibson, Christophe, & Piantadosi, 2017; Y. Sun & Poeppel, 2023). This perspective aligns with the idea that language processing prioritizes structural efficiency rather than raw frequency-based heuristics.

Conclusions. These results provide evidence for the feasibility of content-agnostic features of internal memory structures as drivers of efficient online segmentation. The pattern of performance of our computational- and algorithmic-level model, with the objective of lexicality and length type frequency ranking first, generalizes across two languages of different typological origin. These findings from theory, modeling, and simulation point to a stable phenomenon that deserves further study.

References

Adolfi, F. (2023). *Computational Meta-Theory in Cognitive Science: a theoretical computer science framework* (PhD thesis, University of Bristol). Retrieved from https://hdl.handle.net/1983/c3702d1d-143c-40cc-987e-f2160ea74ac3

Adolfi, F., Bowers, J. S., & Poeppel, D. (2023, May). Successes and critical failures of neural networks in capturing human-like speech recognition. *Neural Networks*, *162*, 199–211. Retrieved 2023-03-13, from https://www.sciencedirect.com/science/article/pii/S0893608023001016 doi: 10.1016/j.neunet.2023.02.032

Adolfi, F., van de Braak, L., & Woensdregt, M. (2024, October). From Empirical Problem-Solving to Theoretical Problem-Finding Perspectives on the Cognitive Sciences. *Computational Brain & Behavior*. Retrieved 2024-10-20, from https://doi.org/10.1007/s42113-024-00216-6 doi: 10.1007/s42113-024-00216-6

Adolfi, F., Wareham, T., & van Rooij, I. (2023). A computational complexity perspective on segmentation as a cognitive subcomputation. *Topics in Cognitive Science*, *15*(2),

255-273.

Beekhuizen, B., Bod, R., & Zuidema, W. (2013). Three design principles of language: The search for parsimony in redundancy. *Language and speech*, *56*(3), 265–290.

Brysbaert, M., & New, B. (2009). Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior research methods*, *41*(4), 977–990.

Cai, Q., & Brysbaert, M. (2010). Subtlex-ch: Chinese word and character frequencies based on film subtitles. *PloS one*, *5*(6), e10729.

Cutler, A. (1994). The perception of rhythm in language. *Cognition*, *50*(1-3), 79–81.

Dautriche, I., Mahowald, K., Gibson, E., Christophe, A., & Piantadosi, S. T. (2017). Words cluster phonetically beyond phonotactic regularities. *Cognition*, *163*, 128–145.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164.

Ferrand, L., Méot, A., Spinelli, E., New, B., Pallier, C., Bonin, P., ... Grainger, J. (2018). Megalex: A megastudy of visual and auditory word recognition. *Behavior Research Methods*, *50*, 1285–1307.

Fortnow, L. (2009). The status of the P versus NP problem. *Communications of the ACM*, *52*(9), 78–86.

Friston, K. J., Sajid, N., Quiroga-Martinez, D. R., Parr, T., Price, C. J., & Holmes, E. (2021). Active listening. *Hearing Research*, *399*, 107998.

Li, J., Bhattasali, S., Zhang, S., Franzluebbers, B., Luh, W.-M., Spreng, R. N., ... Hale, J. (2022). Le petit prince multilingual naturalistic fmri corpus. *Scientific data*, *9*(1), 530.

Perruchet, P., & Vinter, A. (1998). Parser: A model for word segmentation. *Journal of memory and language*, *39*(2), 246–263.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, *21*(6), 322–334.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of verbal learning and verbal behavior*, *9*(5), 487–494.

Sober, E. (1981). The principle of parsimony. *The British Journal for the Philosophy of Science*, *32*(2), 145–156.

Sun, C. C., Hendrix, P., Ma, J., & Baayen, R. H. (2018). Chinese lexical database (cld) a large-scale lexical database for simplified mandarin chinese. *Behavior Research Methods*, *50*, 2606–2629.

Sun, Y., & Poeppel, D. (2023). Syllables and their beginnings have a special role in the mental lexicon. *Proceedings of the National Academy of Sciences*, *120*(36), e2215710120.

Swingley, D., & Algayres, R. (2024). Computational modeling of the segmentation of sentence stimuli from an infant word-finding study. *Cognitive Science*, *48*(3), e13427.

Wareham, H. T. (1996). The role of parameterized computational complexity theory in cognitive modeling. *AAAI-96 Workshop Working Notes: Computational Cognitive Modeling: Source of the Power*.

Zipf, G. K. (2016). *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio books.