

Text Segmentation is Subjective: Empirical Evidence of Human Variability and its Implications for Downstream Tasks

Sophia Conrad
University of Zurich
sophia.conrad@uzh.ch

Florian Frenken
RWTH Aachen University
florian.frenken@ia.rwth-aachen.de

Abstract

Text segmentation, a crucial pre-processing step to many NLP tasks, is often treated as having an objective ground truth, typically approximated by paragraphs. However, this authorial structure may not always reflect how different readers segment the same text. We present an empirical study examining human variation in English text segmentation across 60 texts from three genres. Fourteen annotators segmented texts based on perceived inherent structure, revealing substantial disagreement ($\alpha = 0.29$; $[0.02, 0.62]$). Mean segment lengths vary across *and* within genres, with unscripted speeches exhibiting particularly high variability (2.8–26 sentences). In the same task LLMs show higher intra-model ($\alpha = 0.742$) but comparable inter-model agreement ($\alpha = 0.514$). These findings indicate that text segmentation is shaped by subjective (or model-specific) biases and thus cannot be adequately represented by a single gold standard, calling for a critical re-examination of its role in downstream NLP tasks.

1 Introduction

Text segmentation is a foundational step in natural language processing (NLP), e.g., in discourse analysis, text summarization, topic modeling, information retrieval, and, more recently, retrieval-augmented generation (RAG). In these applications, segmentation is often treated as if there were a single objectively correct solution, ignoring the variability inherent in human segmentation. In this paper, we use RAG as an example downstream application that depends on text segmentation.

RAG enhances the capabilities of large language models (LLMs) by providing them with information outside their pre-training data via access to external knowledge bases (Lewis et al., 2020). This requires documents to be chunked to (1) preserve sufficient context for query matching without breaking up or truncating semantically coherent units,

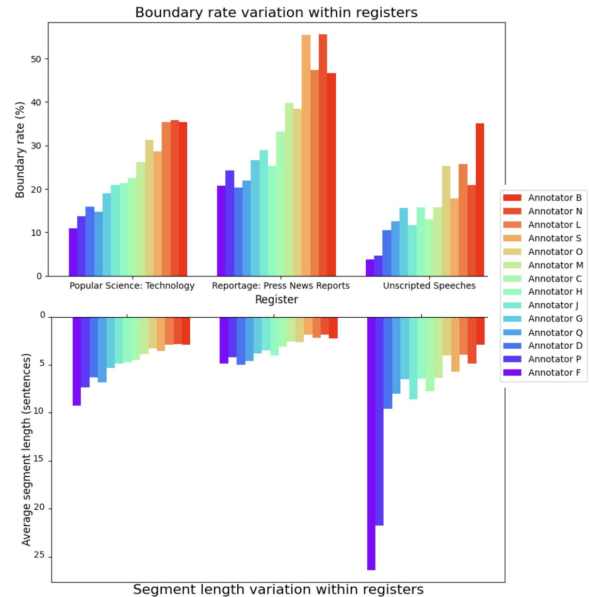


Figure 1: Ranges of boundary rates (top) and ranges of segment lengths (bottom) of each annotator averaged over the texts they annotated.

while (2) avoiding unrelated content as much as possible.

The most common chunking method for RAG currently splits text into fixed-sized token sequences (usually powers of 2 such as 128, 256, or 512) (Gao et al., 2024, p. 8). However, chunk size has been shown to significantly impact retrieval effectiveness (Bhat et al., 2025), suggesting that *how* the text is segmented plays a crucial role. In a systematic comparison, paragraph group chunking specifically – where each chunk consists of multiple adjacent paragraphs – achieves the highest overall retrieval quality based on LLM-as-a-judge evaluation (Shaukat et al., 2026). As such, approaches that preserve “logical” units often outperform fixed-sized chunking methods, motivating paragraph-aware approaches that respect the author’s intended structure (see e.g. Yepes et al., 2024).

Authors typically use paragraphs to organize content into coherent units, structure the development of arguments, and guide readers through the progression of ideas. At the same time, paragraphing practices may be shaped by formulaic norms, such as the five-paragraph essay taught in English-speaking educational contexts (e.g. Tardy, 2019). As a result, paragraphs are best understood as formal artifacts that reflect writing conventions, editorial decisions, and formatting constraints, rather than a uniquely determined underlying structure. While LLMs are trained on large collections of paragraph-structured text such as web pages, they typically observe only one “canonical” segmentation per text, which is often reinforced by duplicate texts with identical structure (Lee et al., 2022, p. 8426). As a result, variation in how a single text could be segmented is not represented; models are only exposed to variation across texts.

This raises the question to what extent segmentations, produced by authors, readers, or LLMs, actually align. In particular, it remains unclear whether paragraph boundaries reflect how readers would segment a text, or whether they serve as convenient but imperfect proxies. To address this gap, we present a large-scale text segmentation study with 10 to 14 annotators per text, showing the variability in human text segmentation. The analysis code is available for replication on GitHub.¹

2 Related Work

Though chunking for the purpose of RAG has the common goal of dividing text into coherent units, text segmentation has typically been treated as a separate task (cf. Nguyen et al., 2025). The focus here has largely been on identifying topical units, starting with Hearst’s (1997) seminal study using patterns of lexical cohesion. When paragraph information is not available, such semantic chunks have also improved information retrieval but are mainly based on embedding similarity (Shaukat et al., 2026). More recent topic modeling approaches have reframed segmentation as a supervised learning task, which may provide further gains in preserving logical coherence. However, for the sake of convenience, they too draw on structural information from paragraphs as training data, in particular from Wikipedia (e.g. Koshorek et al., 2018; Aumiller et al., 2021). We might therefore assume that they likewise suffer from the lack of within-

text variability and produce unnaturally uniform segmentations.

Especially in earlier work studying text structure, single-annotator segmentation was common, typically carried out by an expert (e.g. Cloran, 1994; Gregory, 2002), overlooking potential variability in annotator segmentation choices. Furthermore, segmentation studies were often limited to a single genre, such as news texts (Hirschberg and Grosz, 1992; Grosz and Hirschberg, 1992), conversations between mothers and their children (Cloran, 1994), or spontaneous narrative monologues (Pasonneau and Litman, 1996).

In multi-annotator settings, majority voting is commonly used to derive a gold standard (e.g. Grosz and Hirschberg, 1992; Hirschberg and Grosz, 1992). While this incorporates multiple judgments, it does not ensure that the resulting segmentation is meaningful because these studies often impose arbitrary restrictions, such as making boundary decisions interdependent (e.g., by disallowing adjacent boundaries). The choice of a >50% threshold is also ultimately arbitrary. For example, Hearst (1994) set the boundary at 4/7 annotators, whereas Pasonneau and Litman (1996) considered 3/7 to be the deciding majority. Moreover, some studies restrict where boundaries can be placed, which can artificially inflate agreement. For instance, Pasonneau and Litman (1996) limits annotators to pre-identified prosodic breakpoints, rather than allowing boundaries between any pair of sentences.

Even though Mann et al. (1992) and Johnson (1985) observe that segmentation is a rather subjective task, several later studies report relatively high inter-annotator agreement. For example, Hearst (1993) reports agreement above 80% (across 7 annotators) and regards majority boundaries as the gold standard to which they compare the output of the TextTiling algorithm.

Although Pasonneau and Litman (1996, p. 169) report “far from perfect agreement,” they subsequently present high percent agreement (Gale et al., 1992). Percent agreement is typically defined as the ratio of observed agreements to all possible agreements. However, Pasonneau and Litman (1996) compute it as agreement with the majority opinion relative to the number of possible agreements with that majority. Because the majority label is, by definition, the most frequent choice, at least half of the annotations will agree with it, yielding values of at least 50% even under substantial disagreement. This inflates agreement estimates and obscures the

¹<https://github.com/quantor-project/humsegvar>

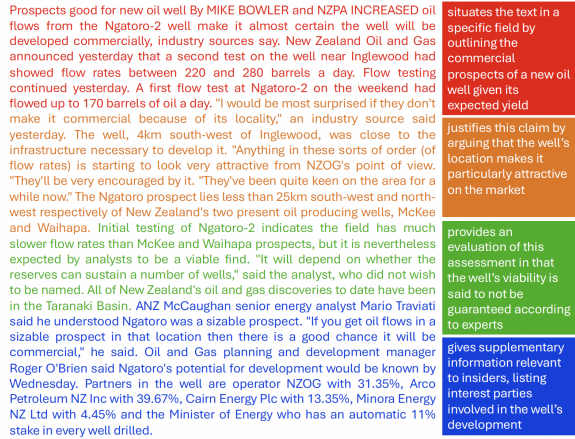


Figure 2: Example from the guidelines with color-coded segmentation of the text (left) and explanations (right).

true extent of annotator disagreement.

3 Methods and Results

We use texts from nine varieties of the International Corpus of English (ICE; Greenbaum, 1996): Canada, Great Britain, Hong Kong, India, Ireland, Jamaica, New Zealand, the Philippines, and Singapore. From these, we randomly select 60 texts² across three different registers: Popular scientific writing in the technology domain, press news reports, and unscripted speeches.

Unlike Passonneau and Litman (1996), who used a corpus of spontaneous narrative monologues, we broaden the range of genres to include both spoken and written texts serving two purposes: First, it allows us to examine whether variability in segmentation persists across a wider range of registers and varieties of English, using the ICE's stratification along both dimensions and thereby strengthening the generality of our findings. Second, the inclusion of written texts enables us to use author-defined paragraphs as an additional point of comparison for annotator decisions.

Following Passonneau and Litman (1996) and Hearst (1994), we asked untrained annotators to segment texts based on an open-ended criterion. While Passonneau and Litman (1996) focused on speaker intention and Hearst (1994) on topic shifts, we instruct annotators to identify segments that make the text recognizable as belonging to a particular genre, i.e., moves (Swales, 1990). We therefore assume that annotators will have an inherent understanding of text structure through repeated

²Note that we use *text* here to refer to each of the *subtexts* in a file (according to ICE parlance) as one self-contained unit.

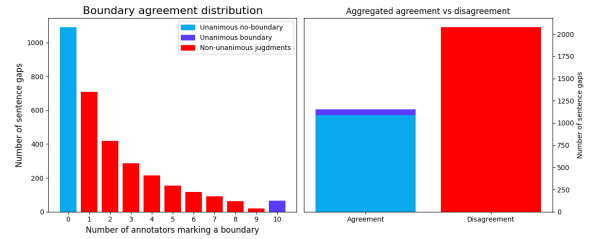


Figure 3: Frequency barplot showing how many annotators (x-axis) perceived a sentence as starting a new segment (left) and aggregated into agreeing and disagreeing cases (right).

exposure in diverse socio-cultural contexts (Halliday and Hasan, 1989). This framing is intended to provide a more intuitive proxy for communicative functions, which should enable more abstract comparisons than segmentations based on a text's specific content.

A tripartite segmentation into introduction, main part, and conclusion was explicitly discouraged, as it would be broadly applicable to texts of any genre. Each segment had to consist of at least one sentence. In addition to the instructions, annotators received an example segmentation (see Fig. 2) that justifies each decision by describing the segment's function in the text. A total of 14 annotators participated in our study, 10 of whom segmented all texts, while the others completed only a subset. The annotators were university students with no background in linguistics. The annotations were performed in the bespoke annotation tool *Segmatator* (Weber and Heinrich, 2026).

To assess the inter-annotator agreement, we compute Krippendorff's α (Krippendorff, 2019), which is suitable for multiple annotators and missing values. We obtain an overall agreement of $\alpha = 0.29$ (see Table 1). Pairwise annotator agreement ranges from 0.02, which is practically chance level, to 0.62, which does not correspond to reliable agreement according to Krippendorff (2019). Fig. 3 shows that annotators largely agree on the absence of boundaries, whereas agreement on boundary placement is sparse. Only a small number of boundaries are consistently identified by all annotators. Applying a majority threshold would therefore retain relatively few boundaries, yielding segments that are substantially longer than author paragraphs.

Fig. 1 shows clear register-level differences: press news reports have the shortest segments ($\mu = 3.3$, $\sigma = 1.1$ sentences) and the highest boundary rate ($\mu = 34.6\%$, $\sigma = 12.6$), while un-

scripted speeches show much longer and more variable segments ($\mu = 8.7$, $\sigma = 6.8$; boundary rate $\mu = 16.3\%$, $\sigma = 8.4$). Popular science texts fall in between ($\mu = 4.8$, $\sigma = 1.9$; boundary rate $\mu = 23.7\%$, $\sigma = 8.5$). Across all registers, we observe substantial variability in both segment length and boundary rate, indicating that segmentation behavior differs markedly not only across text types but also across annotators.

A qualitative analysis of the differences between segmentations reveals that the use of cohesive devices, especially in sentence-initial position, plays a central role in how readers interpret a text structurally (see Appendix A for a concrete example). More specifically, whether they align with the author seems to vary not only with the kinds of textual signals they pick up on but also how this information is weighted relative to each other. For example, lexical chains establish a very obvious link between adjacent sentences but may be superseded when used in combination with a fronted adjunct.

In addition to the manual segmentations, we use GPT-5.1, GPT-4o, and GPT-4o-mini (accessed via the OpenAI API) to segment the same texts with identical instructions and report the overall intra- and inter-model Krippendorff’s α (see Table 1). The example from the guidelines is used to define the raw-text input and desired JSON-formatted output. We mirror the human annotation setup by prompting each model 14 times. To ensure comparability across conditions (28 runs for pairwise comparisons and 42 runs for three-model comparisons), inter-model α is computed using bootstrapping. Specifically, we sample an equal number of runs per model (7 for pairwise comparisons and 5 for three-model comparisons) and calculate α over 500 random draws. We report the mean α and standard deviation across these samples.

The results show relatively high internal agreement across all models compared to human segmentations, with GPT-5.1 achieving the highest consistency (0.782), followed by GPT-4.1 (0.761) and GPT-4o (0.683). In contrast, inter-model agreement is lower, clustering around 0.5, with the strongest alignment between GPT-5.1 and GPT-4o (0.538 ± 0.007). As a result, a substantial portion of variability arises from differences between models. Still, inter-model agreement lies within the upper range of pairwise annotator agreement. Overall agreement across all 15 runs is lower (0.438 ± 0.006), indicating reduced consistency when aggregating across models and runs.

Setting	α
Human annotators (10–14)	0.290
<i>Intra-model agreement</i>	
GPT-5.1 (14 runs)	0.782
GPT-4o (14 runs)	0.683
GPT-4.1 (14 runs)	0.761
<i>Inter-model agreement</i>	
GPT-5.1 vs GPT-4o	0.538 ± 0.007
GPT-5.1 vs GPT-4.1	0.506 ± 0.006
GPT-4o vs GPT-4.1	0.499 ± 0.006
Overall (15 runs)	0.438 ± 0.006

Table 1: Krippendorff’s α for intra-model and inter-model agreement over text segmentations.

4 Conclusion and Outlook

We have shown that inter-annotator agreement on text segmentation is low, indicating substantial variability. In other words, without visual cues such as line breaks, the author’s signals of boundaries are not necessarily recognized. In LLM-generated segmentations, intra-model agreement is consistently high, meaning that individual models produce relatively stable segmentations across repeated runs under identical conditions. In contrast, cross-model agreement is somewhat lower, suggesting that different models impose systematically different structural interpretations on the same text. Exposing models to multiple segmentations per text may increase variability in learned representations, although the extent to which this captures genuinely distinct interpretations rather than surface-level variation remains unclear.

Future studies should investigate whether incorporating within-text variation could be beneficial for downstream tasks. Coming back to RAG as an example, providing an LLM with multiple segmentations of the same text may improve alignment with user expectations regarding which passages are considered informative, especially given evidence that retrieval performance can improve when combining multiple chunking strategies (Yepes et al., 2024, p. 10). Such segmentations could be generated dynamically at the sentence level via prompting approaches (for paragraphs, cf. Duarte et al., 2024) and subsequently aggregated using overlap-based heuristics or majority voting (see mixture of chunkers; Zhao et al., 2025). This would shift the notion of a single “ground-truth” segmentation toward a more consensus-based approximation of relevance.

Limitations

While we cover multiple English varieties and three genres, future work should examine whether the observed patterns generalize to a broader range of texts and whether certain genres exhibit higher levels of agreement than others. In particular, highly structured and conventionalized forms, such as academic research articles, may display more consistent segmentation patterns.

Reliable author-defined segmentation (i.e., paragraph boundaries) was available only for a subset of the texts, which limited the qualitative comparison. Extending this analysis to a larger set of such texts would allow for a more systematic assessment to what extent intuitive human segmentations align with this reference structure.

In an informal exit interview with a subset of annotators, participants reported that the segmentation task was initially challenging but became easier as they grew familiar with recurring structural patterns. However, some texts were perceived as incomplete, which made segmentation more difficult in those cases. This reflects inconsistencies in corpus composition across different varieties of English in the ICE.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 528467412 and the Swiss National Science Foundation (SNSF) Grant (218836)³. We thank the 14 annotators who provided segmentations of the texts.

References

- Dennis Aumiller, Satya Almasian, Sebastian Lackner, and Michael Gertz. 2021. [Structural text segmentation of legal documents](#). In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, pages 2–11, São Paulo Brazil. ACM.
- Sinchana Ramakanth Bhat, Max Rudat, Jannis Spiekermann, and Nicolas Flores-Herr. 2025. [Rethinking Chunk Size For Long-Document Retrieval: A Multi-Dataset Analysis](#). *arXiv preprint*. Version Number: 2.
- Carmel Cloran. 1994. *Rhetorical units and decontextualisation: An enquiry into some relations of context, meaning and grammar*. PhD thesis, Nottingham University.
- André V. Duarte, João Marques, Miguel Graça, Miguel Freire, Lei Li, and Arlindo L. Oliveira. 2024. [LumberChunker: Long-Form Narrative Document Segmentation](#). *arXiv preprint*. ArXiv:2406.17526 [cs.CL].
- William Gale, Kenneth Ward Church, and David Yarowsky. 1992. [Estimating upper and lower bounds on the performance of word-sense disambiguation programs](#). In *Proceedings of the 30th annual meeting on Association for Computational Linguistics -*, pages 249–256, Newark, Delaware. Association for Computational Linguistics.
- Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. 2024. [Retrieval-Augmented Generation for Large Language Models: A Survey](#). *arXiv preprint*. ArXiv:2312.10997 [cs.CL].
- Sidney Greenbaum, editor. 1996. *Comparing English Worldwide: The International Corpus of English*. Clarendon Press, Oxford.
- Michael Gregory. 2002. Phasal analysis within communicative linguistics: two contrastive discourses. In Peter H. Fries, Michael Cummings, David Lockwood, and William Spruiell, editors, *Relations and functions within and around language*, Open Linguistics, pages 316–345. Continuum Press, London.
- Barbara Grosz and Julia Hirschberg. 1992. [Some intonational characteristics of discourse structure](#). In *2nd International Conference on Spoken Language Processing (ICSLP 1992)*, pages 429–432.
- M.A.K. Halliday and Ruqaiya Hasan. 1989. *Language, context, and text: aspects of language in a social-semiotic perspective*. Oxford University Press, Oxford.
- Marti A. Hearst. 1993. TextTiling: A quantitative approach to discourse segmentation. Technical report, Technical report, University of California, Berkeley, Sequoia.
- Marti A. Hearst. 1994. [Multi-paragraph segmentation expository text](#). In *32nd Annual Meeting of the Association for Computational Linguistics*, pages 9–16, Las Cruces, New Mexico, USA. Association for Computational Linguistics.
- Marti A Hearst. 1997. [TextTiling: Segmenting text into multi-paragraph subtopic passages](#). *Computational linguistics*, 23(1):33–64.
- Julia Hirschberg and Barbara Grosz. 1992. [Intonational features of local and global discourse structure](#). In *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, February 23-26, 1992*.
- Nancy S. Johnson. 1985. Coding and analyzing experimental protocols. *Handbook of discourse analysis*, 2.

³<https://data.snf.ch/grants/grant/218836>

- Omri Koshorek, Adir Cohen, Noam Mor, Michael Rotman, and Jonathan Berant. 2018. [Text Segmentation as a Supervised Learning Task](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 469–473, New Orleans, Louisiana. Association for Computational Linguistics.
- Klaus Krippendorff. 2019. *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, Inc.
- Katherine Lee, Daphne Ippolito, Andrew Nystrom, Chiyuan Zhang, Douglas Eck, Chris Callison-Burch, and Nicholas Carlini. 2022. [Deduplicating training data makes language models better](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8424–8445, Dublin, Ireland. Association for Computational Linguistics.
- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS ’20*, Red Hook, NY, USA. Curran Associates Inc.
- William C. Mann, Christian M.I.M. Matthiessen, and Sandra A. Thompson. 1992. [Rhetorical Structure Theory and Text Analysis](#). In William C. Mann and Sandra A. Thompson, editors, *Pragmatics & Beyond New Series*, volume 16, page 39. John Benjamins Publishing Company, Amsterdam.
- Hai Toan Nguyen, Tien Dat Nguyen, and Viet Ha Nguyen. 2025. [Enhancing Retrieval Augmented Generation with Hierarchical Text Segmentation Chunking](#). *arXiv preprint*. Version Number: 1.
- Rebecca J. Passonneau and Diane J. Litman. 1996. [Empirical analysis of three dimensions of spoken discourse: Segmentation, coherence, and linguistic devices](#). In *Computational and conversational discourse: Burning issues—An interdisciplinary account*, pages 161–194. Springer.
- Muhammad Arslan Shaukat, Muntasir Adnan, and Carlos C. N. Kuhn. 2026. [A Systematic Investigation of Document Chunking Strategies and Embedding Sensitivity](#). *arXiv preprint*. Version Number: 1.
- John M. Swales. 1990. *Genre analysis*. Cambridge University Press.
- Christine M. Tardy. 2019. Is the five-paragraph essay a genre? *Changing practices for the L2 writing classroom: Moving beyond the five-paragraph essay*, pages 24–41.
- Timm Weber and Philipp Heinrich. 2026. Segmatator. <https://github.com/fau-klue/segmatator>.
- Antonio Jimeno Yepes, Yao You, Jan Milczek, Sebastian Laverde, and Renyu Li. 2024. [Financial Report Chunking for Effective Retrieval Augmented Generation](#). *arXiv preprint*. Version Number: 3.
- Jihao Zhao, Zhiyuan Ji, Zhaoxin Fan, Hanyu Wang, Simin Niu, Bo Tang, Feiyu Xiong, and Zhiyu Li. 2025. [MoC: Mixtures of Text Chunking Learners for Retrieval-Augmented Generation System](#). *arXiv preprint*. Version Number: 2.

A Qualitative Analysis

To compare annotator segmentations with author paragraph structure in more detail, we analyze one example text (see Appendix B), focusing on *where* boundaries differ and potential reasons *why*. These findings are broadly representative of the kinds of differences that can be found between author paragraphs and reader segmentations. Crucially, neither interpretation is necessarily ‘correct’; it rather illustrates how text structure can be perceived very differently depending on what one attends to.

In general, there are many sentences where all annotators agreed on the absence of a boundary. These are usually related to how the text develops a coherent argument through the use of (especially sentence-initial) cohesive devices. For instance, the start of sentence 25 “One customer’s hotline” picks up on a specific example of the “customer hotlines” mentioned in sentence 24. And sentence 26 then posits a contrast with “Other hotlines”. This would explain why 5 annotators identified a boundary after sentence 26 as the focus switches to “How promptly complaints are dealt with”.

However, if we interpret this sequence not linearly but functionally as an evaluation of “customer service quality”, it makes sense why the author would include that sentence in the paragraph as well. This ambiguity might then also explain the disagreement between author and readers from sentence 27 onward. For example, even though sentence 28 explicitly mentions *quality*, its placement inside the initial circumstantial adjuncts *From a qualitative standpoint* linguistically suggests a structural break.

B Example Text

ID	Text	# Ann.
1	Quality takes a knock in a liberalised telecom industry	
		11
2	SIMON CHAN	
		4
3	January 1 was a milestone in the history of the telecommunications industry in Hong Kong.	
4	With the opening up of local fixed-line telecom network services, Hong Kong has liberalised the telecoms industry, from carriers to valued-added service providers.	2
		3
5	The good news is that users can enjoy a wide range of services from different operators at a competitive price.	
6	But there has also been an unprecedented number of complaints .	1
		5
7	What has gone wrong ?	
8	Is the quality of service really degraded?	
9	What are the expectations for quality of service?	
10	And what should we do to fix it?	
		7
11	A common myth is that competition will improve quality.	
11	Not true.	
		7
12	The complaint statistics from the Office of the Telecommunications Authority (Ofa) show an alarming trend.	
		1
13	The number of complaints is projected to rise 36 per cent from 1,733 in 2001 to 2,361 for last year.	
14	In 1997, there were 87 complaints.	
		2
15	Recent statistics from the Consumer Council revealed 4,684 cases in 2001, up 51 per cent on 2000.	
		2
16	Of these complaints , Internet services have overtaken mobile-phone services as the top troublemakers.	
		7
17	It is also obvious that from a business consumer's perspective, the quality of service criteria for different services are not transparent.	
18	Some of these have been crafted in Service Level Agreements (SLA), but the majority are never published.	
		2
19	Even in the recent network jam incident of September 11 last year, the lack of transparency between fixed and mobile networks led a lot of people to complain about the quality of our networks, even though the latter are among the world's best.	
		3
20	The fixed network is engineered to allow a maximum of one out of 100 calls to be blocked , and five out of 100 for the mobile network except in emergency circumstances.	
		1
21	These network quality-of-service measurements should be published regularly so that consumers have more information when choosing an operator.	
		4
22	Other performance measurements include time to install new services; network reliability, repeated customer fault reports, the ability to meet targeted repair times, complaint handling and billing.	
23	These statistics may have been available to the individual operators, but were never made public for the obvious reasons that they might hurt the carriers' image.	
		9
24	A classic example of poor customer service is the maze of machine voices on customer hotlines when you try to contact a human for assistance .	
25	One operator's customer hotline is unmanned and, after forcing the anxious customer to wait in vain for up to seven minutes, it drops the call.	
26	Other hotlines will hide the operator behind several voice menus so that only their staff can reach them .	
		5
27	How promptly complaints are dealt with is a good indicator of quality of service and can be used to select a new service.	
		2
28	From a qualitative standpoint, customer-satisfaction surveys can be conducted to either match those described previously or supplement other aspects such as customer satisfaction.	
		3
29	In order to reap good quality in a fully liberalised market such as Hong Kong's, we should not just set our sights on lowering prices, quality must be made transparent.	
		4
30	The danger is that we let quality fall below a minimum acceptable level, affecting our reputation as a telecommunications hub of Asia.	
		3
31	For example , Internet surfing response times to overseas Web sites will degrade for second-tier ISPs as they drive down costs by adding new users without increasing their bandwidth .	
		1
32	Data on quality will have to come from each operator.	
33	Will they provide such sensitive data?	
34	It will not be easy.	
		1
35	This is where the regulator can help.	
		4
36	In Britain , the telecom regulator Ofel published a set of Comparable Performance Indicators (at www.cpi.org.uk) to help consumers compare the quality of services from a range of British telecom companies.	
		1
37	In Hong Kong , Ofa is in the best position to publish similar indicators , so that we can truly become a quality telecom hub of Asia.	
		11
38	Only then will consumers , operators and Hong Kong as a whole have the chance to win.	
		8
39	Simon Chan is the chairman of the Hong Kong Telecommunications Users Group, representing business telecom users to the SAR government.	
		1
40	The group 's Web site is at www.hktug.org .	