The prosody of Clefted Relatives: A new window into prosodic representations

Buhan Guo[1], *Nino Grillo*[1], *Sven Mattys*[1], *Andrea Santi*[2], *Shayne Sloggett*[1], *Giuseppina Turco*[3]

[1]University of York
[2]University College London
[3]Laboratoire de Linguistique Formelle, UMR 7110, CNRS/Université Paris Cité

buhan.guo@york.ac.uk, nino.grillo@york.ac.uk, sven.mattys@york.ac.uk, a.santi@ucl.ac.uk, shayne.sloggett@york.ac.uk, giuseppina.turco@cnrs.fr

Abstract

The well-attested association between information structure and the acoustic properties of sentences can be captured by either assuming a direct mapping between semantics and acoustics or invoking the mediation of phonological processes operating on well-defined prosodic domains (indirect approaches). Although these two accounts' predictions typically converge, we identified an understudied contrast for which the two views make different predictions. Specifically, through 3 experiments (1 production, 2 comprehension), we tested the prosody of it-clefts containing string-identical Connected Clauses (*-Who sang? -It was [the editor] [that sang]*) or Relative Clauses (*-Who called? -It was [the editor [that sang]] ([that called])*) that have semantically focused elements of different structural sizes. Connected Clauses attach high in the structure and are given. Relative Clauses are assumed to convey background information, but here they are nested within the focused element and also in focus. Our production results showed a localized prominence on the rightmost stressable syllable of the Relative Clause, which is in line with indirect accounts. The comprehension studies further showed that i) clefted Relatives trigger garden-path effects in reading, but ii) garden-paths disappear when prosody is present. The studies support indirect accounts by employing more complicated structural configurations.

Index Terms: prosodic structures, information structure, sentence processing, prosodic disambiguation, syntax-prosody interface

1. Introduction

Within the domain of information structure, meaning differences map onto well-recognized acoustic differences. Whether this mapping is *direct* [1, 2, 3] or mediated by linguistic/prosodic representations ([4, 5], a.m.o) is, however, a contentious matter. The two families of accounts have been hard to differentiate because their predictions typically align. The examples in (1) and (2) provide a good illustration of this issue in the domain of information structure. Both accounts in fact predict words carrying *new* information to be associated with more prominent acoustic features (including longer duration, higher intensity and wider pitch range) than words which are *given* or carry lower informational load. Thus, the word *pizza* tends to be accented in answer to the Question (1) but not to the Question (2), and the opposite obviously holds for the word *John*. This is because for the Question (1), *pizza* is new information and in focus, while for (2) it is associated with background information and marked as given.

(1) - What does John like?
- John likes PIZZA.

(2) - Who likes pizza?
- JOHN likes pizza.

However, when focus falls on a single word, as in the example above, it is impossible to decide whether information structure directly determines the prominence of that word or whether this is mediated by intermediate phonological representations (e.g. a *+accent* feature associated with focused elements in a phonological representation). More sophisticated work on narrow vs. broad focus using similar SVO structures [6] provides clear evidence on the nature of prosodic correlates of focus, but still does not distinguish between the two accounts.

We argue that these accounts do in fact make different predictions and that these predictions can be tested when looking at focused elements with more complex internal structures than the single word examples above, e.g. complex Noun Phrases (NP) as in (3). Since each part of the complex NP *pizza with anchovies* in (3) carries new information, direct accounts should predict higher prominence for the whole phrase. Indirect accounts, however, make very specific and localized predictions about accent assignment, which are relatively independent from the informational content of a word and are determined on the basis of rules which apply to syntactic and phonological representations (e.g. the Nuclear Stress Rule [7, 8]).

(3) What does John like?
John likes [$_{NP}$ pizza [$_{PP}$ with anchovies]].

To test this proposal, we investigated the prosodic properties of clefted Noun Phrases modified by restrictive Relative Clauses (RC), as in (5) and compared them with string-identical sentences in which only the initial Noun Phrase was clefted (4). As clarified by the context questions and the bracketing, while the two structures are string identical, they display very different syntactic and semantic properties. The prototypical cleft sentence in (4) involves focus on the clefted subject *the humorist*, with the Complementizer Phrase (CP) *that was leaving the scene* being a *Connected Clause* (CC), introducing given information and being obligatorily extraposed and linked to the matrix clause [9]. The example in (5), on the other hand, involves focalization of a complex NP which also contains a nested Relative Clause (*the humorist that was leaving the scene*). In this example, such interpretation is ensured by mismatching the content of the context questions and that of the CP, which excludes a Connected Clause reading.

(4) - Who was leaving the scene?
- It was [$_{NP}$ *the humorist*] [$_{CC}$ that was leaving the scene].

(5) -Who called?
- It was [$_{NP}$ *the humorist* [$_{RC}$ that was leaving the scene]] ([$_{CC}$ *that called*]).

Using string identical sentences with distinct structural and interpretive properties via clefted elements of different structural sizes has the advantage of leading to different predictions for the *direct* and *indirect* accounts across the two structures. While the two accounts make similar predictions for prototypical clefts like in (4), i.e. they both predict prominence to fall solely on the clefted simple Noun Phrase *the humorist* (this is supported by [10]), their predictions may differ in the Relative Clause condition as in (5). Since the whole Complex NP is clefted and carries new information, direct accounts should predict a generalized higher prominence across the whole phrase, including each region of Relative Clauses. Indirect accounts, however, would predict highly localized effects of accent placement when a Complex Noun Phrase containing a Relative Clause is clefted: Focus stress should fall on the most deeply nested word, i.e. *scene* in (5).

One additional reason to investigate the role of prosody in the disambiguation of string identical sentences like in (4) and (5) is that clefted Relative Clauses have been recently shown to generate garden path effects in the absence of prosody [11]. This obviously raises the question of whether explicit prosody can disambiguate these two readings and avoid a garden path effect, as observed in reading clefted Relatives.

We investigate the predictions of the direct and indirect accounts in two experiments, a planned production study and an auditory perception study in English.

2. Experiment 1: Planned Production

2.1. Participants

Five native British English speakers (3 women) originating from different regions of the UK participated in the experiment in a soundproof booth (age range=24-to-35, age average=29.8, SD=4.6). Participants gave their informed consent and were paid for their participation. Each subject participated in the experiment twice with at least a one-week gap between sessions to ensure that each session focused on a single critical structure.

2.2. Materials

Each condition contained 24 Question-Answer pairs and each answer was structured as follows: It was + the NP1 + that was + Verb + the NP2 (as shown in examples (6) and (7)). Structures in focus were marked in *italics*. Stimuli were prosodically controlled across items, keeping the number of syllables and the position of lexical stress constant within each region.

(6) CONNECTED CLAUSE CONDITION:
- Who was leaving the scene?
- It was [$_{NP}$ *the humorist*] [$_{CP}$ that was leaving the scene].

(7) RELATIVE CLAUSE CONDITION:
- Which one of them was identified?
- It was [$_{NP}$ *the humorist* [$_{CP}$ *that was leaving the scene*]].

The experimental items were interspersed with 48 fillers that included varied syntactic structures and matched experimental items in length. Twelve fillers were also preceded by questions to make half of all items form Question-Answer pairs.

2.3. Procedure

Participants were instructed to silently scan the entire (question and) sentence before reading aloud and then produce the questions (if any) and sentences naturally and fluently at normal speed. Items were automatically presented on a computer screen and recorded on a PC using the software ProRec 2.4 (©Mark Huckvale, University College London).

Experimental stimuli were initially divided into two lists to ensure that each participant only produced one critical structure in each session. All items were pseudo-randomised, such that another two lists were made whose items were respectively the same as their original version but presented in the reversed order to avoid potential sequence effects, leading to a total of four lists. Experimental items were separated by at least one filler item in every list.

Every session started with four practice items, followed by 24 experimental items interspersed with 48 fillers, leading to a total of 76 items for each participant in each session. The whole experiment lasted approximately 40 minutes.

2.4. Statistical Analysis

Segmentation was performed automatically using the Montreal Forced Aligner [12]. Duration, F0 and intensity were automatically detected using scripts ran in Praat software [13]. The results of the automatic procedure were checked and manually corrected (blinded to the condition the sentence belonged to) in case of errors.

Based on the previous discussion about the predictions from the two accounts, for each sentence, we selected regions with lexical words for comparison (NP1 *humorist*, Verb *leaving*, and NP2 *scene*). Particularly, to test for localized effects, we focused on comparing NP2 vs. Verb, in which we expect to see the most different predictions from the two accounts. Within these regions, we extracted and analysed the following three measurements: 1) raw duration in ms, 2) F0 range in semitones, and 3) raw intensity in dB. Statistical analysis was performed using linear mixed-effects regression models in the R-package lme4 [14] with a maximum structure. Likelihood ratio tests (LRTs) were used to examine the significant contribution of fixed effects to the model. Post-hoc analyses were performed using package *emmeans* to further test for structural effects at each region. For every model, we set the fixed effects to Region (*NP2* vs. NP1 vs. Verb), Structure (*CC* vs. RC), and the interaction between them. The element in *italics* indicated the reference level.

2.5. Results

2.5.1. Duration

Localised effects of Structure were found for duration as presented in Table 1 and Figure 1. Model comparison showed significant interactions between Region and Structure (AIC = 7609.0, $\chi 2(2) = 95.81$, $p < .001$). Crucially, the difference in structural effects at Verb was significantly smaller than at NP2 (β = -33.75, SE = 7.58, $t(701)$ = -4.45, $p < .001$), indicating localized effects as predicted by the indirect account. This is further supported by Post-hoc analyses which revealed that Connected Clauses had significantly shorter duration than in Relative Clauses at NP2 (β = -48.6, SE = 9.16, $t(701)$ = -5.30, p =.004), but not at NP1 (p =.008) or Verb (p =.15).

Table 1: *Average raw duration in ms (with SD in parentheses)*

	NP1	**Verb**	**NP2**
CC	479(*82.3*)	308(*42.9*)	359(*77.7*)
RC	455(*81.7*)	323(*46.8*)	407(*62.0*)

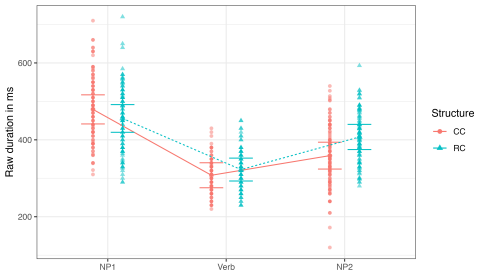

Figure 1: *Duration (in ms) at NP1, Verb and NP2*

2.5.2. *F0 Range*

For F0 range, we observed a similar localised pattern to duration, illustrated in Table 2 and Figure 2. Models revealed significant effects of interactions (AIC = 2747.1, $\chi 2(2) = 28.82$, $p < .001$) and more importantly, significantly localised effects at NP2 compared to Verb (β = -2.40, SE = 0.46, $t(606)$ = -5.23, $p < .001$). This is supported by the significantly smaller F0 range for Connected Clauses than Relative Clauses at NP2 (β = -1.88, SE = 0.35, $t(606)$ = -5.34, $p < .001$), but not at NP1 (p =.16) or Verb (p =.90).

Table 2: *Average F0 range in semitones (st) (with SD in parentheses)*

	NP1	Verb	NP2
CC	4.62(*2.90*)	2.27(*2.70*)	1.46(*1.75*)
RC	4.10(*2.35*)	2.24(*1.54*)	3.55(*2.41*)

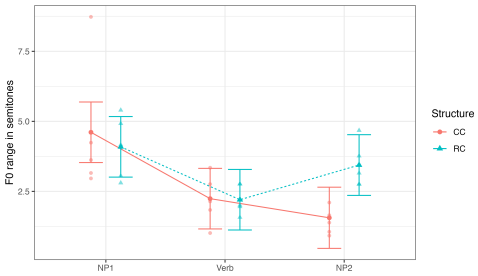

Figure 2: *F0 range (in st.) at NP1, Verb and NP2*

2.5.3. *Mean Intensity*

Similarly, localized effects in intensity were observed as shown by Table 3 and Figure 3. Models showed significant interactions between Structure and Region (AIC = 3229.6, $\chi 2(2) = 29.13$, $p < .001$) with significantly different structural effects at NP2 compared to at Verb (β = -1.44, SE = 0.35, $t(698)$ = -4.10, $p < .001$). Nevertheless, at NP2, Connected Clauses were produced with only numerically lower intensity than Relative Clauses (β = -4.31, SE = 1.19, $t(698)$ = -3.64, p =.06), indicating a weaker effect compared to duration and F0 range.

2.6. Intermediate Discussion

Experiment 1 established that (at least some) speakers prosodically disambiguate Relative Clauses and Connected Clauses, as evidenced by differences in duration, F0, and intensity. These differences, in particular duration and F0 range, appear to be highly localized, in line with predictions of indirect accounts. In Experiment 2 we test whether listeners are sensitive to these acoustic differences and can use them to assist syntactic processing and avoid garden path effects with Relative Clauses observed in the absence of prosody in [11].

Table 3: *Average intensity in dB (with SD in parentheses)*

	NP1	Verb	NP2
CC	62.2(*4.57*)	59.3(*3.94*)	54.5(*4.51*)
RC	64.8(*3.62*)	62.3(*4.02*)	58.9(*4.16*)

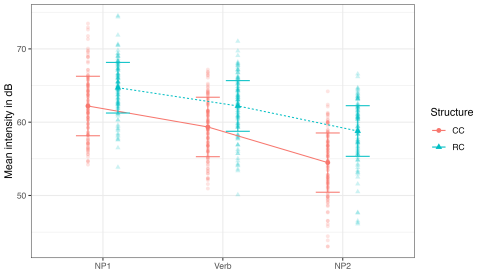

Figure 3: *Intensity (in dB) at NP1, Verb, and NP2*

3. Experiment 2: Auditory Perception

3.1. Participants

Sixty-four native speakers of English (30 women) located in the US (mean age=33.8, SD=8.1) were recruited via the online recruitment platform Prolific (www.prolific.com). We used recruitment filters to ensure that all participants had no language, vision or hearing-related disorders.

3.2. Materials

In the current study, each of the 24 experimental stimuli comprised two parts: a preceding context (including a question like in (6) and (7)) plus an audio stimulus as an answer to the question. The 24 audio stimuli were sentences in Experiment 1, produced by a trained linguist with either a Connected Clause or Relative Clause prosody following the patterns in Experiment 1. Context and Questions together either elicit a Connected Clause or Relative Clause reading of the answer. Taken together, this experiment has a 2 Context (CC vs. RC leading) * 2 Prosody (Matched vs. Mismatched prosody of the recording to the context) design. Experimental items were balanced for conditions and were interspersed with 36 fillers, preceded by 3 practice items. In total, each participant completed 63 trials.

To ensure participants' attentiveness, half of the experimental items and fillers contained comprehension questions that targeted different parts of the context to avoid strategic reading of the context. The proportion of *Yes* and *No* answers to the comprehension questions was balanced.

3.3. Procedure

This experiment followed a paradigm in [10] and was performed on the Gorilla Experiment Builder (www.gorilla.sc). In each trial, after reading the context and question, participants

listened to a recording with either a matched or mismatched prosody to the given context. Next, they were asked to judge whether they thought the audio sentence was an acceptable answer for the context and question by choosing *Yes* or *No*. Every judgment was followed by a confidence rating, asking about their certainty in that judgment (*Not confident*, *Somewhat confident*, or *Very confident*). Finally, participants needed to answer a comprehension question, if any.

3.4. Data Analysis

Trials with a Reaction Time of less than 200 ms or longer than 10000 ms in the acceptability judgement were excluded from analysis, leading to a total of 52 items being removed, accounting for 3.38% of the data. We analysed both raw binary data from acceptability judgement (*Yes* or *No*) and the responses combining both the binary acceptability data with graded confidence ratings. The combined data resulted in a 6-point scale, ranging from 1-*Very confident unacceptable* to 6-*Very confident acceptable*, which resembled the Likert scale. Due to space limits, only the 6-point data were reported here. Considering the ordinal nature of the data, statistical analysis was performed using cumulative link mixed-effects models (CLMM) in the ordinal package [15]. Context, Prosody, and their interactions were set as fixed factors with a maximum random effects structure while allowing for model convergence. Package emmeans was used in Post-hoc analysis to examine the simple effect of matched and mismatched prosody under different contexts.

3.5. Results

Figure 4 presented the distribution of the 6-point rating across conditions. Mismatched Prosody to the Context received significantly lower ratings than the Matched one (β = -0.76, SE = 0.21, z = -3.62, p <.001), while no significant difference between Contexts was found (p =.77). However, and more interestingly, our data illustrated a significant interaction between Prosody and Context on the ratings: Mismatched Prosody was rated much lower for the RC-leading Context, compared to the CC Context (β = -1.70, SE = 0.33, z = -5.20, p <.001). Interestingly, Post-hoc analysis showed that mismatched prosody was associated with higher acceptability for CC-leading context than RC ones (β = 1.68, SE = 0.35, z = 4.83, p <.001), but such difference was not found for matched prosody (p =.45), indicating participants' preference for the two structures became consistent with the help of a matched prosody.

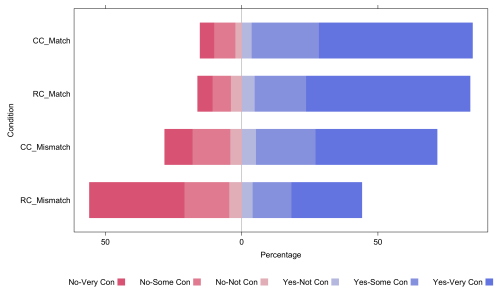

Figure 4: *Distribution of the 6-point ratings*

4. Discussion and Conclusions

This study demonstrates prosodic disambiguation for the previously untested contrast between string-identical Connected Clauses vs. Relative Clauses in both production and perception. In Experiment 1, speakers prosodically disambiguated between the two structures through diverse patterns of duration, F0 range and intensity. Prosodic disambiguation was also observed for perception in Experiment 2, where listeners showed sensitivity to the differences in prosodic features between the structures. Moreover, although clefted Relative Clauses lead to garden path effects in the absence of prosody [11], the preference for Connected Clauses disappeared when the target prosody was provided, suggesting that cooperative prosody eliminated garden path effects. That prosodic disambiguation appeared to be more beneficial for Relative Clauses than Connected Clauses further supports Guo et al.'s [11] findings that clefted Relative Clauses are more difficult to process than Connected Clauses.

More importantly, we argued that testing narrow focus on complex Noun Phrases containing nested Relative Clauses can shed light on the mapping between information structure and surface prosodic patterns of sentences. The localized effects shown in Experiment 1 (and the sensitivity to these localized effects displayed by listeners in Experiment 2) provide further evidence for the existence of intermediate, linguistic levels of representation between sound and meaning. This is specifically shown by a mismatch in the structural effects for NP2 vs. Verb across duration, pitch, and (to a somewhat lesser extent) intensity: a larger effect of structure on the measurements was observed for NP2 relative to the other regions of interest. Such patterns align with the predictions from the indirect account as discussed earlier, and are less compatible with the direct view which would expect these two regions to show parallel structural effects. This argument, i.e. that one account allows more specific predictions on the localization of an effect, echoes previous work at word level by [16], who showed that word-level lengthening is influenced by domain-edge effects in ways predicted by phonological accounts but not by direct approaches. Linguistic principles governing focal accent assignment [8] indeed make even more specific predictions about the localisation of the effect, which we aim to test in future work.

We presented a case for employing more sophisticated syntactic configurations (and in particular structural nesting) when investigating the prosodic realization of information structure. The primary benefit of this approach is to provide multiple regions of interest for prosodic analysis, which enables testing the predictions put forth by both direct and indirect accounts in more detail. In future research, our objective is to delve deeper into clefted Relative Clauses, aiming to differentiate the relative contributions of various prosodic variables in tracking constituent structure and information structure (see also [17]).

In conclusion, our results show that the prosodic pattern of focused constituents appears to be governed by specific principles (e.g., the Nuclear Stress Rule) [4, 5, 7, 8] which make reference to linguistic levels of representation. These results, i.e. the localized effects of focus on prosody, are more in line with predictions put forth by indirect accounts.

5. Acknowledgements

This work was jointly funded by the University of York (Psycholinguistics PhD Grant 2022-2025) and the Laboratoire de Linguistique Formelle, the French Investissements d'Avenir-Labex EFL program (ANR-10-LABX-0083), contributing to the IdEx Université Paris Cité - ANR-18-IDEX-0001.

6. References

[1] W. E. Cooper, S. J. Eady, and P. R. Mueller, "Acoustical aspects of contrastive stress in question–answer contexts," *The Journal of the Acoustical Society of America*, vol. 77, no. 6, pp. 2142–2156, 1985.

[2] P. Lieberman, "Some effects of semantic and grammatical context on the production and perception of speech," *Language and Speech*, vol. 6, no. 3, pp. 172–187, 1963.

[3] Y. Xu, A. Lee, S. Prom-on, and F. Liu, "Explaining the penta model: a reply to arvaniti and ladd," *Phonology*, vol. 32, no. 3, p. 505–535, 2015.

[4] D. R. Ladd, *Intonational Phonology*, 2nd ed., ser. Cambridge Studies in Linguistics. Cambridge University Press, 2008.

[5] J. B. Pierrehumbert, "The phonology and phonetics of english intonation," Ph.D. dissertation, Massachusetts Institute of Technology, 1980.

[6] M. Breen, E. Fedorenko, M. Wagner, and E. Gibson, "Acoustic correlates of information structure," *Language and Cognitive Processes*, vol. 25, no. 7, pp. 1044–1098, 2010.

[7] N. Chomsky and M. Halle, *The sound pattern of English*. Harper Row, 1968.

[8] M. L. Zubizarreta, "165Nuclear Stress and Information Structure," in *The Oxford Handbook of Information Structure*. Oxford University Press, 2016.

[9] M. Reeve, "Clefts," Ph.D. dissertation, University College London, 2010.

[10] A. Arnhold, "Prosodic focus marking in clefts and syntactically unmarked equivalents: Prosody–syntax trade-off or additive effects?" *The Journal of the Acoustical Society of America*, vol. 149, no. 3, pp. 1390–1399, 2021.

[11] B. Guo, N. Grillo, S. Mattys, A. Santi, S. Sloggett, and G. Turco, "Prosody Disambiguates String-Identical Connected Clauses and Relative Clauses," in *AMLaP29 (Architectures and Mechanisms of Language Processing 2023)*, 2023, p. 397.

[12] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Proc. Interspeech 2017*, 2017, pp. 498–502.

[13] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot. Int.*, vol. 5, no. 9, pp. 341–345, 2001.

[14] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[15] R. H. B. Christensen, *ordinal—Regression Models for Ordinal Data*, 2023, r package version 2023.12-4.

[16] L. White and A. E. Turk, "English words on the procrustean bed: Polysyllabic shortening reconsidered," *Journal of Phonetics*, vol. 38, no. 3, pp. 459–471, 2010.

[17] M. Wagner and M. McAuliffe, "The effect of focus prominence on phrasing," *Journal of Phonetics*, vol. 77, p. 100930, 2019.