Prosodic focus marking in clefts and syntactically unmarked equivalents: Prosody–syntax trade-off or additive effects?

Anja Arnhold[a]

Department of Linguistics, University of Alberta, Edmonton, Alberta T6G 2E7, Canada

ABSTRACT:

Two experiments quantitatively investigated the interaction of prosody and syntax in marking focus in English. A production study with 28 participants (analyzing 919 utterances) found that the acoustic marking of subject focus vs broad focus, induced through a preceding context question, was generally the same in clefts as in sentences with unmarked syntax. Thus, results suggested that prosody is independent from syntax rather than showing a trade-off (weaker prosodic marking for clefts). Focus was marked with f0 range, f0 maxima, f0 minima, duration, and intensity. Maxima of focused subjects were not significantly higher, but they were earlier than in broad focus. In a perception experiment, 230 participants rated the suitability of 24 auditorily presented stimuli as answers to preceding context questions inducing subject focus or broad focus. Clefts and sentences prosodically marking the subject as focused were rated higher in subject focus than in broad focus contexts. Syntax and prosody did not interact, again suggesting the absence of a trade-off. Thus, both studies suggest an additive use of syntax and prosody: Prosodic focus marking was equally extensive and effective in the presence of syntactic focus marking as without.
© 2021 Acoustical Society of America. https://doi.org/10.1121/10.0003594

I. INTRODUCTION

One of the most important functions of speech prosody in English—and many other languages—is the marking of information structure, i.e., how the content of an utterance relates to the context of knowledge, beliefs, and assumptions shared between speaker and addressee [e.g., see Krifka (2008) for an overview of information structural concepts and Kügler and Calhoun (2020) for a cross-linguistic overview of prosodic marking]. An important facet of information structure is the division of utterances into focused and backgrounded parts, with the general assumption that speakers make the focused part prosodically prominent. Thus, when the sentence *The doorman was moving a mirror* answers the question *Who was moving a mirror?* the focused *the doorman* will be prosodically marked as prominent, but when it answers *What was the doorman moving?* the focus and thus the prominent part is *a mirror*.

However, even though it has long been recognized that information structure is crucial in determining the prosodic shape of English utterances (e.g., Gussenhoven, 1983; Halliday, 1967), English also has other, morpho-syntactic, ways of marking information structure. This article focuses on cleft sentences like *It was the doorman who was moving a mirror* and how their use interacts with prosodic focus marking. In particular, it asks whether the two means of marking focus—prosody and clefting—truly interact or whether they are simply two independent linguistic means available to speakers.

Several studies have investigated the acoustic correlates of focus in English (Breen *et al.*, 2010; Cooper *et al.*, 1985; Eady and Cooper, 1986; Eady *et al.*, 1986; Pell, 2001; Sánchez-Alvarado, 2020; Wagner and McAuliffe, 2019; Xu and Xu, 2005). The comprehensive study by Breen *et al.* (2010) tested various acoustic measures with a carefully designed quantitative analysis. Consistent with previous research, they found that focused words were reliably distinguished from pre- and post-focal words in the same utterance by higher maximum intensity, higher mean and maximum fundamental frequency (f0), and longer durations. In addition to these instances of narrow focus where only a part of the utterance—subject, verb, or object—was focused, they also studied broad focus, where the whole utterance is focused, answering questions like *What happened?* Comparing narrow object focus to broad focus, they found that maximum f0, duration, and maximum intensity were more uniform across the sentence in broad focus than in object focus and that objects had higher intensity, higher f0, and longer durations in narrow focus than in broad focus, with intensity contributing most to distinguishing the two conditions.

For clefts, it is usually assumed that the nuclear—i.e., most prominent—accent appears on the clefted constituent, i.e., *the lawyer* in *It was the lawyer who was mailing the letter*, at least for prototypical ("specificational") clefts as employed in the present studies (Declerck, 1988; Hedberg, 1990; Rochemont, 1986). However, the only empirical studies of the prosody of English clefts seem to be the corpus studies by Collins (1991, 2006), who evaluated prosodic annotations and confirmed that the majority of clefted

[a]Electronic mail: arnhold@ualberta.ca

constituents in it-clefts indeed carried nuclear accents, mostly with falling pitch. The acoustic characteristics of English clefts have not, as far as the author is aware, been studied experimentally.

It is also commonly assumed that clefts are equivalent in terms of information structure to sentences with unmarked syntax and a nuclear accent on the same constituent, i.e., that *It was the owner who rewarded the waiter* and *The OWner rewarded the waiter* (capitals indicate nuclear accent) both mark *the owner* as focus (Atlas and Levinson, 1981; Lambrecht, 2001).[1] Accordingly, cross-linguistic prosody–syntax interactions in terms of categorical grammatical choices have been identified, such as languages with less prosodic flexibility making greater use of clefts (e.g., Lambrecht, 2001).

This article investigates a different kind of interaction by looking at cases within one language where both prosody and clefts could be used, alternatively or together. In such a situation, it asks whether prosody and syntax show trade-offs or are used additively. In the production of prosodic focus marking, its quantitative interactions with syntax—i.e., the prosodic differences in the simultaneous presence vs absence of marked syntax—have mostly been studied for languages with flexible word order (e.g., Calhoun, 2015; Luchkina, 2016; Patil *et al.*, 2008; Skopeteas *et al.*, 2009). This research has presented evidence that in such languages, word order modulates the perception of prominence when prosody is kept constant (Vainio and Järvikivi, 2006), speakers adjust prosody to compensate for a marked word order in a broad focus context (Vainio and Järvikivi, 2007), and, perhaps most pertinently, prosodic focus marking is less extensive when constituent order can be adjusted than when the task requires speakers to use only the unmarked word order (Arnhold and Féry, 2013). The present article investigates whether such effects can also be found in English, where word order is generally fixed.

For the perception of clefts in English, Calhoun *et al.* (2019) showed that participants follow the syntactic cue when syntax and prosody conflict, e.g., perceiving subject focus for a subject cleft with the nuclear accent on the object. They did not, however, compare utterances with congruent prosodic and syntactic focus cues to utterances with either only prosody or only clefts marking focus. This will be the approach in the present article, which asks directly whether the perceived appropriateness of focus marking further increases when both prosody and clefting are employed additively.

II. PRODUCTION EXPERIMENT

The production study was designed to answer the following research question: Do prosodic and syntactic focus marking interact, such that prosodic focus marking differs in cleft sentences compared to unmarked syntax?

A. Materials and methods

Native speakers of Canadian English were asked to produce clefts and sentences with unmarked syntax in response to narrow subject focus questions, as well as a broad focus baseline.

1. Participants

A total of 39 undergraduate students participated in the study. Data from 11 participants were discarded for the following reasons: They identified a language other than English as their primary language and/or the first language they acquired (8), speech was not recorded due to experimenter error (2), or they were distracted by nearby construction noise (1). Thus, data were analyzed from 28 participants who had all grown up in Canada (21 female, 7 male; age: 18–43 yrs, mean: 22.43 yrs, standard deviation: 5.59 yrs).

The study was approved by the Research Ethics Board 2 of the University of Alberta.

2. Materials

The experimental materials contained 24 target sentences, which appeared in two syntactic conditions: with unmarked syntax, i.e., as simple subject-verb-object (SVO) sentences, and in a cleft construction. Each constituent (S, V, and O) had three syllables, with the middle syllable stressed, and consisted of sonorant segments as much as possible.

The target sentences appeared in two different contexts, which induced either broad focus or narrow focus on the subject. For broad focus conditions, the target sentence on a whole constituted the answer to a preceding question like *What happened?, What was that about?*, etc., whereas for subject focus conditions, the preceding context ended in a question starting with *Who*, so that the subject constituent was the part of the target sentence that answered this question (e.g., *Who was wearing a jewel? — Maria was wearing a jewel*). The full list of sentences, an example context, and further details appear in the supplementary materials.[2]

The factors syntax and information structure were not fully crossed because clefts are not well-formed in broad focus. Thus, there were three experimental conditions: (1) broad focus and unmarked syntax, (2) subject focus and unmarked syntax, and (3) subject focus and cleft.

The experimental materials (24 items × 3 conditions = 72 target sentences with contexts) were combined with filler trials of a similar structure and distributed onto four lists so that each list consisted of 51 trials total (36 targets and 15 fillers), which appeared in two blocks.

3. Procedure

Each session consisted of instructions and informed consent, two practice trials, the two experimental blocks with an optional pause between them, and finally, the participants completing a background questionnaire.

During each trial, participants first read the context silently. At the end of each context, the target sentence appeared in blue font, and participants spoke it out loud as an answer to the directly preceding question. To keep

J. Acoust. Soc. Am. **149** (3), March 2021

Anja Arnhold 1391

participants alert, they were asked a question about the context they had just read after 12 trials on each list, with a choice of two possible answers. Participants whose data were evaluated on average answered 93.75% of these questions correctly.

Participants were recorded in a sound-attenuated booth with a Countryman (Menlo Park, CA) headset microphone (H6 Omni) placed about 4 cm from their mouths and a Fostex (Tokyo, Japan) field recorder (model FR-2LE) at 16-bit resolution and a sampling frequency of 44,100 Hz.

4. Data editing and analysis

Of 1008 target trials (28 participants × 36 trials), the recorder was not switched on for 6 utterances, and a further 83 (8.2%) were discarded due to mispronunciations, slips of the tongue, or other disfluencies. Thus, 919 utterances were analyzed. They were manually segmented in PRAAT (Boersma and Weenink, 2020) based on visual and auditory inspection, using silent intervals, fricative noise, the third formant, and the second formant to guide segmentation, in this order, where available. Manually edited pitch objects were used to measure f0, with measurement errors like octave jumps removed and the first couple of cycles at the beginning and end of voiced intervals trimmed to avoid microprosodic influences. Since acoustic measures were obtained separately for all subject, verb, and object constituents, 2757 data points were analyzed for each dependent variable.

The following dependent variables were modeled individually with linear mixed-effects models, using the package lme4 in R (Baayen et al., 2008; Bates et al., 2015; R Core Team, 2020): f0 range (f0 maximum – f0 minimum; in semitones, st), f0 maximum (in st relative to a reference frequency of 100 Hz), f0 minimum (in st relative to a reference frequency of 100 Hz), time of f0 maximum relative to the beginning of the stressed (i.e., second) vowel (in ms), constituent duration (in ms), and mean intensity measured over the center 50% duration of the stressed vowel (in dB, scaled to a reference level of 50 dB).

Model selection always started with a model containing an interaction between the predictors condition (levels: broad focus and unmarked syntax, subject focus and unmarked syntax, subject focus and cleft) and constituent (subject, verb, object), as well as random effects. For models of f0 measures, participant gender was also included as a control predictor. The best-fitting model for each measure was determined through model comparisons such that the model was only as complex as justified by the improved fit to the data (Matuschek et al., 2017); see supplementary materials for details on the procedure.[2]

The package lmerTest (Kuznetsova et al., 2017) was used to obtain p-values in fixed effect summaries. To identify significant differences between individual predictor levels in models with interactions, pairwise comparisons were conducted with the lsmeans function from the package emmeans (Lenth, 2020).

B. Results

All acoustic measures revealed a significant interaction between condition and constituent, indicating that conditions influenced the prosodic realization of subjects, verbs, and objects in different ways. Sections II B 1–6 detail significant differences as indicated by pairwise comparisons calculated based on these models. Fixed effects summaries of all best models and pairwise comparisons based on them appear in the supplementary materials.[2] In addition to showing effects of the manipulated variables, the best linear mixed-effects models of the height of f0 maxima and f0 minima of constituents contained a significant effect of gender, in line with male speakers producing overall lower f0 values than female speakers, which is expected and thus not further discussed below.

Before Secs. II B 1–6 detail the results of statistical analyses of individual measures by condition and constituent, Fig. 1 illustrates time-normalized average f0 contours across sentences by condition. These are based on measurements of f0 at ten equidistant points for each constituent. To facilitate comparison between the conditions, no measurements are displayed for
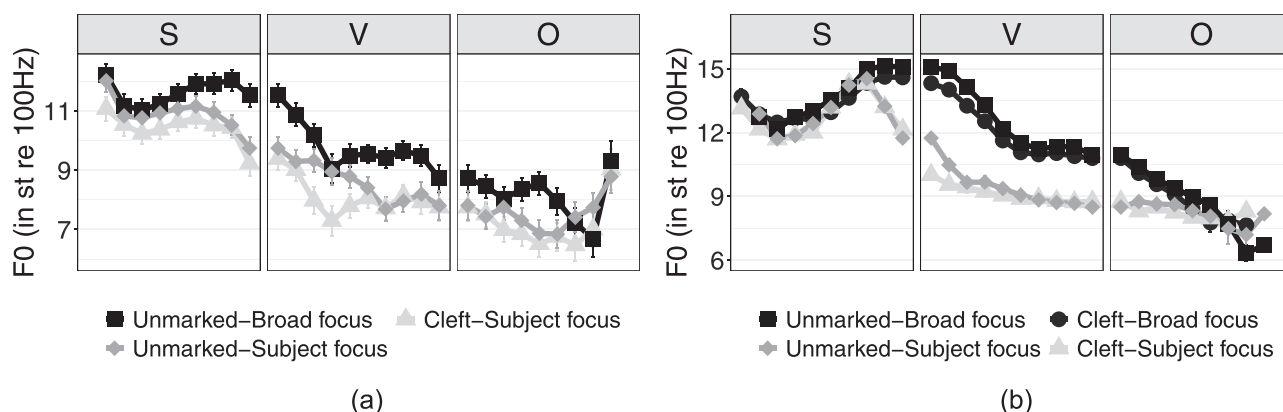

FIG. 1. Average time-normalized f0 contours for subject (S), verb (V), and object (O) constituents in different conditions. (a) Results of production study. (b) Stimuli for perception study. Error bars indicate standard errors.

linguistic material appearing only in cleft condition (i.e., *It was/is* and *who*).

Figure 1(a) is based on the data of the production study discussed here. Measurements from the first point for subjects and the last point for objects were trimmed because of missing values and variation in measurable f0. Apart from this variation, the average contours as produced by the participants are similar to the stimuli used for the perception study in Fig. 1(b), discussed further below. In both cases, focused subjects did not show higher f0 maxima compared to broad focus, in contrast to the average contours shown in Xu and Xu (2005) and the measurements reported in Breen *et al.* (2010). Instead, participants lowered f0 on the

post-focal constituents compared to broad focus. This lowering began with a clear fall starting already on the focused subject. Importantly, contours were very similar in the two subject focus conditions.

1. f0 range

Subject constituents had a significantly larger f0 range in both subject focus conditions compared to broad focus, while the difference between focused subjects in clefts and in sentences with unmarked syntax was not significant, as indicated by pairwise comparisons; see Fig. 2(a) for illustration. For both verbs and objects, f0 range was significantly
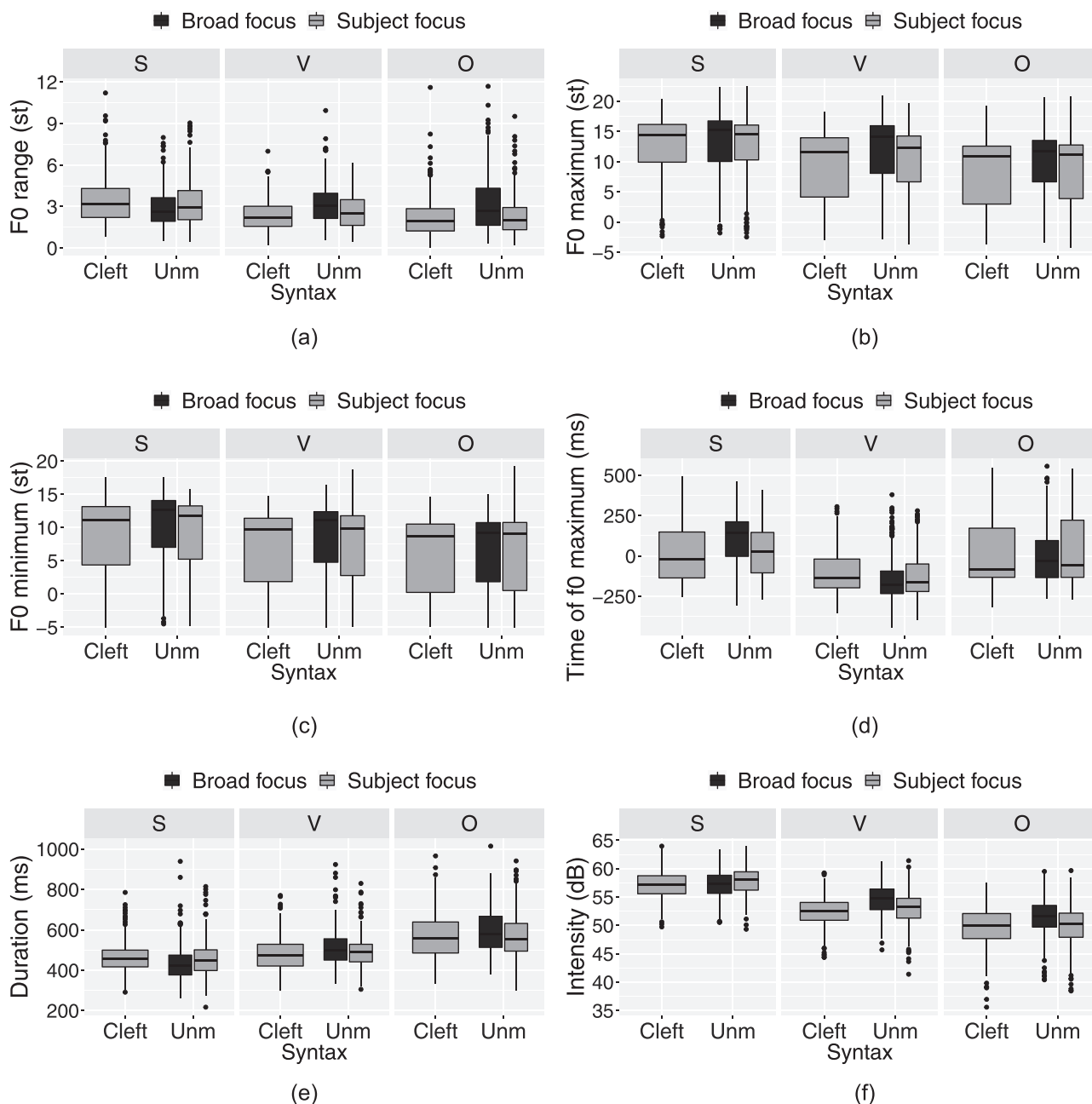

FIG. 2. Distribution of acoustic measurements by condition (subject focus and cleft, subject focus and Unm = unmarked syntax, broad focus and Unm = unmarked syntax) and constituent (S = subject, V = verb, O = object). (a) f0 range; (b) f0 maximum; (c) f0 minimum; (d) time of f0 maximum; (e) constituent duration; (f) mean intensity of stressed vowel.

smaller in the two subject focus conditions compared to broad focus. For verbs, ranges were additionally significantly smaller in clefts than in sentences with subject focus and unmarked syntax.

Regarding differences between the constituents, subjects had significantly larger ranges than both verbs and objects in the two subject focus conditions, while verbs and objects did not significantly differ from each other. The broad focus condition did not show significant differences between any pair of constituents.

2. f0 maximum

The height of f0 maxima on subjects did not differ significantly between conditions, although maxima were marginally higher in subject focus with clefts than in broad focus [see Fig. 2(b)]. By contrast, f0 maxima of verbs and objects were significantly lower in both subject focus conditions than in broad focus. For verbs, maxima were also significantly lower in clefts than in subject focus sentences with unmarked syntax. Comparing constituents, subjects had significantly higher maxima than both verbs and objects in all conditions. Also, the maxima of verbs were higher than those of objects in all conditions. This is in line with a general downtrend of f0 over the course of sentences.

3. f0 minimum

f0 minima of subjects were lower in both subject focus conditions than in broad focus, with no significant differences between the two subject focus conditions [Fig. 2(c)]. The same was true of verbs, in line with the fact that f0 falls starting on subjects usually continued into the verb constituent (see Fig. 1). Object minima did not differ significantly between conditions but were marginally lower in clefts than in broad focus. In line with a general pitch downtrend over the course of each sentence, minima of subjects were higher than those of verbs and objects, and minima of verbs were higher than those of objects in all conditions.

4. Alignment of f0 maximum

f0 maxima were significantly earlier for subjects in both subject focus conditions than in broad focus, with no significant difference between the two subject focus conditions [Fig. 2(d)]. Verbs and objects did not differ significantly between the conditions, except that maxima of verbs were significantly later in clefts compared to broad focus. Maxima of verbs in all three conditions were often at the very beginning of the verbal constituent, which was spanned by an f0 fall from the maximum of the subject [see Fig. 1(a)]. Accordingly, verbs had significantly earlier maxima than both subjects and objects in all three conditions. Finally, in broad focus, maxima of subjects were marginally later than those of objects.

To allow direct comparison with the results of Eady and Cooper (1986) and Breen et al. (2010), the location of the f0 maximum was also calculated as a proportion of the duration of the constituent. Results were the same as with the absolute measure, except that the difference between subjects and objects in broad focus was significant, whereas the differences between verbs in cleft condition and verbs in broad focus, between subjects and verbs in cleft condition, and between verbs and objects in cleft condition did not reach significance.

5. Constituent duration

Subject constituents had longer durations in subject focus conditions than in broad focus, with no significant difference between the two subject focus conditions [Fig. 2(c)]. Conversely, verbs and objects had shorter durations in both subject focus conditions than in broad focus. Additionally, verbs had shorter durations in clefts than in subject focus sentences with unmarked syntax. Differences between the constituents show a consecutive lengthening over the course of the sentence so that in broad focus, verbs had significantly longer durations than subjects, while objects had longer durations than both subjects and verbs. In the subject focus conditions, however, only the differences between verbs and objects and between subjects and objects were significant, whereas the elongated subject constituents were not significantly shorter than the verbs.

6. Intensity

Intensity in the stressed vowels of subject constituents was significantly higher in subject focus sentences with unmarked syntax than in clefts and in broad focus, while subject intensity did not significantly differ between clefts and broad focus [Fig. 2(f)]. For both verbs and objects, intensity was significantly reduced in the two subject focus conditions compared to broad focus. Additionally, verbs had significantly lower intensity in clefts than in subject focus sentences with unmarked syntax. Moreover, intensity showed a significant downtrend over the course of sentences in all three conditions: Values were significantly higher for subjects than for verbs and objects and additionally significantly higher for verbs than for objects.

When intensity was measured as the maximum reached during the whole constituent instead, the measure found to best distinguish conditions in the study by Breen et al. (2010), the same significant differences appeared.

C. Interim summary and discussion

This study asked whether prosodic and syntactic focus marking interact, such that prosodic focus marking differs in cleft sentences compared to those with unmarked syntax. Results consistently showed the same prosodic focus marking in both syntactic conditions. This was true for those strategies increasing the prominence of the focused subject constituent itself and for those reducing the prominence of the other two, backgrounded, constituents (with one possible exception in the use of intensity; see below). Therefore, from a qualitative point of view, it can be concluded that prosodic focus marking is the same whether or not it is combined with syntactic focus marking in the form of clefting.

This leaves the possibility of a quantitative interaction, i.e., an adjustment in the extent of prosodic focus marking. Such an adjustment should be detected in significant differences between the two subject focus conditions. For example, subject focus in sentences with unmarked syntax could be expected to lead to a larger f0 range on the subject than when subject focus is already marked by clefting—even if both conditions had larger subject f0 ranges than broad focus. Only one such difference appeared: Intensity of the focused subject constituent was significantly lower for clefts than for unmarked syntax. This could be evidence for a trade-off between prosodic and syntactic focus marking, but another explanation is possible and, given the other findings, more likely: Subjects in cleft sentences probably had lower intensity than those in subject focus with unmarked syntax because they were preceded by other linguistic material instead of being sentence-initial. This fits with the general pattern of intensity declining over the course of a sentence [also see Wagner and McAuliffe (2019)]. It also matches the fact that verbs, like subjects, showed lower intensity in clefts than for subject focus with unmarked syntax.

A few other differences between clefts and unmarked syntax appeared for verb constituents but were not in line with a trade-off. Generally, the two subject focus conditions differed significantly from broad focus, but not from each other. In sum, the present results do not support the assumption of a trade-off relationship between prosodic and syntactic focus marking in English.

A further aspect deserving discussion is the nature of the observed prosodic focus marking strategies. In line with previous research by Breen et al. (2010) and others, participants realized focused subjects with larger f0 ranges, lower f0 minima, longer durations, and—though less clearly—increased intensity compared to broad focus, while at the same time producing backgrounded constituents with smaller f0 ranges, lower f0 maxima, shorter durations, and lower intensity than in broad focus. However, in contrast to findings by Breen et al. (2010), the speakers participating in this study did not realize focused subjects with raised f0 maxima. Further, again unlike speakers analyzed by Breen et al. (2010), they produced focused subjects with significantly earlier f0 maxima. Thus, instead of expanding the f0 range of the focused constituent by raising the maximum and thus expanding both the rise to the peak and the following fall, they expanded only the f0 fall by starting it earlier than in broad focus, reaching a lower minimum at the end.

This finding is, however, not as unexpected as it may first appear. Other studies have failed to find raised f0 peaks for focused sentence-initial subjects (Cooper et al., 1985; Eady and Cooper, 1986; Sánchez-Alvarado, 2020). Also, Breen et al. (2010) only directly compared broad and narrow focus for object constituents, whereas f0 peaks of focused subjects were compared to the peaks of backgrounded constituents. However, Eady et al. (1986) and Xu and Xu (2005) do find higher peaks in initial position in a direct comparison with broad focus. Various potential explanations for these discrepant findings have been

suggested in passing, including an influence of sentence length [Eady et al. (1986); but see the results in Pell (2001)] and a difference between clause-initial and utterance-initial position (Wagner and McAuliffe, 2019).

The present results support the suggestion by Eady and Cooper (1986) and Eady et al. (1986) that in sentence-initial position, the fall from the peak and the subsequent lowering of f0 on post-focal constituents are more important cues to focus than raised peaks [also see Sánchez-Alvarado (2020)]. Like the present study, Eady and Cooper (1986) and Eady et al. (1986) also found that sentence-initial subjects had earlier f0 peaks when they were in narrow focus than they did in broad focus or when they were part of the background while another word was focused. As earlier peaks extend the pitch fall and/or the following lower pitch component, the absence of raised peaks and the earlier timing of peaks can be understood as part of the same prosodic focus marking strategy. Note that while most studies do not evaluate the timing/alignment of peaks, Breen et al. (2010) and Xu and Xu (2005) consistently find raised peaks, but not consistently earlier f0 maxima. Further research is needed to establish what factors determine which of the two strategies is used to mark focus via f0 in sentence-initial position.

III. PERCEPTION EXPERIMENT

A perception study was designed to address the following research question: Do prosodic focus marking and clefting interact in conveying information structure?

Based on existing research, it is expected that subject cleft sentences will be more acceptable in contexts that indicate subject focus than in broad focus contexts. Also, prosody marking the subject as focused should be more acceptable in subject focus contexts than in broad focus contexts. In addition to empirically testing these predictions, the present study aims to find out whether these effects are additive. For example, in a subject focus context, the acceptability of a sentence with prosody marking the subject as focused could be further increased by clefting. Alternatively, syntax and prosody could show a trade-off relationship such that, for example, prosodic marking has a smaller effect on perception when combined with simultaneous clefting.

A. Materials and methods

Participants heard clefts and syntactically unmarked sentences with varying prosody and were asked to rate how well they fit broad focus and subject focus contexts.

1. Participants

A total of 300 undergraduate students participated. For 60 of them, English was not both their first and their primary language, while 10 additional participants spoke a variety of English other than Canadian English or did not specify which variety they spoke. Their data was thus discarded, leaving 230 participants for analysis (169 female, 59 male, 2

no response; age: 17–45 yrs, mean: 20.5 yrs, standard deviation: 3.2 yrs)

The study was approved by the Research Ethics Board 2 of the University of Alberta.

2. Materials

The 24 items from the production study were used, again appearing in two syntactic conditions, as a cleft and with unmarked syntax. This syntactic manipulation was crossed with a two-level prosodic variable: A female native speaker of Canadian English in her early twenties read all stimulus sentences in two versions, once following the context eliciting broad focus and once following the context eliciting subject focus as used in the production experiment. Thus, she produced one version each that was prosodically marked as broad focus and one that was prosodically marked as subject focus. The speaker was asked to make the prosodic information structure marking very clear at the beginning of the recording session. She repeated the first few items several times, until she and the author agreed that she had produced distinct realizations suitable for the respective contexts. Note that while in the production study, participants were not asked to produce cleft sentences in broad focus because of the unnaturalness of this combination; the speaker did record this combination for the perception study.

As visible in Fig. 1(b), the speaker's intonation clearly distinguished the two prosodic conditions but, within those, was nearly identical for the two syntactic conditions. Importantly, for the three conditions also appearing in the production study, her intonation was very similar to the average contours of the production experiment, though naturally it showed less variation. In particular, while the broad focus and subject focus conditions were audibly clearly distinct, like the participants in the production study, the speaker did not raise the f0 maximum of the subject in subject focus, but rather produced an earlier fall compared to broad focus.

Participants encountered all four versions of these spoken stimuli in two different written contexts, which set up an expectation of either broad focus or subject focus on the target sentence, i.e., the spoken stimulus. These contexts were the same as used in the production experiment and as used by the speaker who produced the spoken stimuli to elicit broad focus and subject focus. Thus, all 24 items occurred in eight conditions: two syntactic conditions (clefts vs unmarked syntax) × two prosodic conditions (marked as broad focus vs subject focus) × two context conditions (broad focus vs subject focus).

The resulting 24 × 8 = 192 target trials were distributed onto eight lists with a Latin square design. In addition to the 24 target trials, each list contained 36 filler trials. The fillers likewise consisted of spoken stimuli presented after contexts ending in a question; see supplementary materials for further details.[2]

Each list thus consisted of 60 trials, which were divided into two blocks, each containing equal numbers of target and filler trials. These were preceded by two practice trials.

3. Procedure

Participants were tested in small groups of up to 12 people in a computer lab. After receiving instructions and giving informed consent, participants moved through the experiment at their own pace and filled out a background questionnaire at the end of the experimental session.

As in the production study, participants read the preceding context paragraph by paragraph. After reading the question at the end of the context, they heard the spoken stimulus via headphones. Next, they were asked to rate how well the answer they heard fit on a scale from 1 ("completely unsuitable") to 7 ("completely suitable"). They were able to listen to the sound multiple times if they wished.

To ensure their attentiveness, participants responded to the same questions about the contexts as in the production study after 16 trials, of which eight were target trials. Due to an error in the experiment file, responses to these questions were not recorded for the first 239 participants. For the 40 analyzed participants with recorded responses, 93.75% were correct on average.

B. Analysis and results

Linear mixed-effects models were fit to the 5520 target responses (230 × 24 items) following the same procedure as described for the production study, but testing the predictors syntax (levels: cleft, unmarked), context (levels: broad focus, subject focus), and prosody (levels: broad focus, subject focus).

Model selection started with a model containing a three-way interaction between the three predictors, but model comparison showed that this three-way interaction did not significantly improve the model's fit to the data. A two-way interaction between syntax and prosody did not improve model fit either and was thus likewise removed. The best model only contained an interaction between context and syntax, as well as a second interaction between context and prosody as predictors; see supplementary materials for fixed effects summary and pairwise comparisons.[2]

Participants generally rated the suitability of the stimulus higher in a context inducing subject focus than following a question inducing broad focus (see Fig. 3). In line with this, both clefts and sentences with unmarked syntax were rated significantly lower in broad focus than in subject focus contexts. Pairwise comparisons further indicated that cleft sentences were rated lower than sentences with unmarked syntax in broad focus contexts, whereas both syntactic conditions received equally high ratings in subject focus contexts, possibly due to a ceiling effect.

Pairwise comparisons for the interaction between context and prosody again showed that answers to broad focus questions were rated overall lower, and this difference was

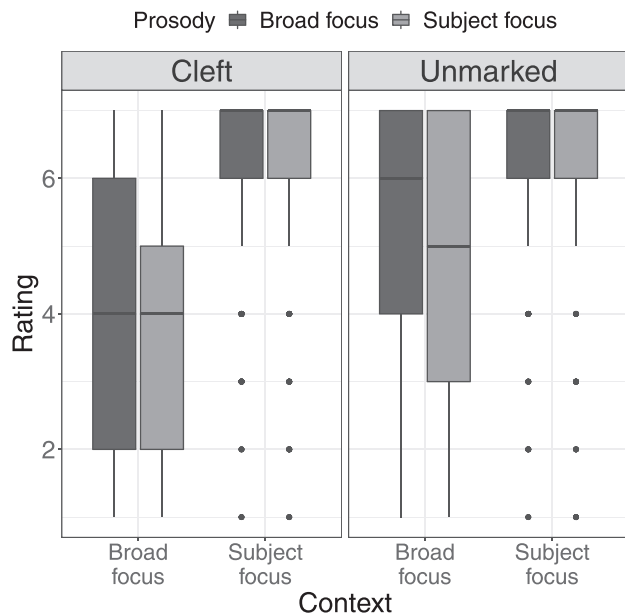1396   J. Acoust. Soc. Am. **149** (3), March 2021

Anja Arnhold

FIG. 3. Acceptability ratings in perception study by context, syntax, and prosody of the stimulus.

significant for answers in both prosody conditions. Further, in broad focus contexts, stimuli with broad focus prosody received higher ratings than those with subject focus prosody. In subject focus contexts, however, the two prosodic realizations did not differ significantly.

C. Discussion

The results of this study did not show a significant interaction between prosody and syntax. Thus, clefts were rated more highly in subject focus than in broad focus contexts, as expected, but this was true independently of their prosodic realization. Conversely, sentences prosodically marking subjects as focused were rated as more suitable in subject focus than broad focus contexts, whether or not they were clefts. Thus, there was no indication of a trade-off between syntax and prosody.

Instead, the results suggested an additive effect: The effects of prosody and syntax were simultaneously present. Thus, cleft sentences received lower ratings than unmarked syntax in broad focus contexts and additionally lower ratings when their prosody also indicated that the subject was focused. Likewise, sentences prosodically marking the subject as focused were less acceptable in broad focus than those without focus-marking prosody, also for clefts [similarly, Kember *et al.* (2019) found an additive positive effect of clefts and prosodic focus marking on memory]. Crucially, the absence of an interaction between syntax and prosody indicated that the size of the effect of prosodic marking did not change significantly when combined with clefting, and vice versa.

This additive effect is, however, not apparent in the subject focus condition (Fig. 3). This could be interpreted as the absence of an additional effect of a second focus marking strategy when subject focus is already marked by either prosody or syntax, in line with the findings by Calhoun *et al.*

(2019), who did not observe an additive effect of prosody and clefting with a methodology differing slightly from the present one.

However, a closer look at the subject focus contexts does not support this interpretation. Instead, the absence of a syntax-prosody-interaction is in line with a ceiling effect. Neither the prosodic nor the syntactic conditions differed significantly from each other in subject focus, and answers in these contexts were generally rated highly. Arguably, narrow focus questions are more specific than broad focus ones, and it is thus possible that participants interpreted this difference as answers with subject focus providing a better fit of the answer to the preceding question. Interestingly, this strong tendency to rate any answer to a subject focus question as inherently appropriate even persisted for those stimuli with prosody suitable for broad focus. Importantly, modeling failed to find a three-way interaction between context, prosody, and clefting, suggesting that the data on a whole do not support an interaction between prosody and syntax, and that the relationship between them was similar in both subject and broad focus contexts.

In sum, while other interactions were significant, the results of the present perception study do not lend support to the hypothesis of a trade-off relationship between prosodic and syntactic focus marking. Instead, they provide some support for an additive relationship, though future studies with a range of methodologies are desirable to strengthen this finding.

IV. GENERAL DISCUSSION

Both the production experiment and the perception study failed to show evidence for a quantitative trade-off between prosodic and syntactic focus marking. Instead, the results were more in line with additive effects. In production, speakers produced essentially the same prosodic focus marking whether they used clefting simultaneously or not. In perception, both syntax and prosody affected the suitability of an utterance as an answer to a preceding broad or subject focus question, but their effects were not reduced (or increased) in the presence of a second means of focus marking.

This suggests that prosodic focus marking is not optional even when subject focus is already marked via clefting. In other words, at least for specificational clefts, prosodic focus marking of the clefted constituent seems to be a necessary part of the form of cleft constructions.

An interesting question is whether these results are specific to English or even to the use of cleft constructions. As reviewed in the Introduction, evidence for prosody–syntax trade-offs has been found for languages with flexible word order. These studies differ from the present one in the type of syntactic variation studied—word order vs clefting. But also, because they regularly employ constituent order to convey information structure, prosody may have a different relative importance in these languages. As has often been observed, English is both relatively rigid with respect to syntax and prosodically very flexible, so that information

J. Acoust. Soc. Am. **149** (3), March 2021

Anja Arnhold   1397

structure exerts a powerful influence on the prosodic shape of an utterance. It therefore makes sense that prosodic focus marking is so central that it cannot be "switched off" even when it is redundant in the presence of a syntactic strategy to mark information structure.

A similar argument about the central importance of prosodic focus marking in English appears in Kember *et al.* (2019). They found that while both for English and for Korean, participants best recalled target words when they were both clefted and prosodically prominent, the languages differed when only one of the linguistic means was employed. Clefting on its own provided a greater memory advantage than prosodic prominence alone for Korean, whereas for English, prosodic marking provided the greater advantage. Kember *et al.* (2019) suggest this difference may be due to the relative frequency with which syntactic focus marking is used in the two languages. One difference between their results for English and those of the present study is that they found clefting alone to not significantly improve memory compared to a baseline where the target was neither clefted nor prosodically prominent. By contrast, clefting clearly influenced the suitability of answers to broad and subject focus questions in the present perception study, independently of prosody. This is in line with the idea that clefting has specific pragmatic functions, including focus marking, but does not generally convey prominence and thus affect memory in the way that prosody does.

V. CONCLUSION

The present results indicate that prosody and clefting are used in an additive way to mark information structure, rather than being in a trade-off relationship. Thus, prosodic prominence of a focused word and, conversely, reduced prominence of the background were the same in production even in the presence of clefts. Prosody also equally influenced the perception of information structure with and without simultaneous clefting. Prosodic prominence of the clefted constituent appears to be an inherent component of specificational cleft constructions.

ACKNOWLEDGMENTS

[1]This is not to say they are identical semantically and/or pragmatically, but all narrow focus contexts used here were suitable for characteristics of clefts, particularly regarding exhaustivity (e.g., Destruel et al., 2015). Also, while several scholars assume that not all clefts have this information structure (e.g., Declerck, 1988), the clefts in the present studies clearly do.

[2]See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0003594 for item list, example context, details of statistical modeling, model summary tables and pairwise comparisons for both studies, and description and analysis of fillers in the perception study.

Arnhold, A., and Féry, C. (**2013**). "The role of syntactic flexibility and prosody in marking given/new distinctions in Finnish," Finno-Ugric Lang. Linguist. **2**(2), 2–32.

Atlas, J. D., and Levinson, S. C. (**1981**). "It-clefts, informativeness and logical form: Radical pragmatics (revised standard version)," in *Radical Pragmatics*, edited by P. Cole (Academic Press, New York), pp. 1–62.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**(4), 390–412.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**, 1–48.

Boersma, P., and Weenink, D. (**2020**). "Praat: Doing phonetics by computer (version 6.1.14) [computer program]," http://www.praat.org (Last viewed 05/15/2020).

Breen, M., Fedorenko, E., Wagner, M., and Gibson, E. (**2010**). "Acoustic correlates of information structure," Lang. Cogn. Proc. **25**, 1044–1098.

Calhoun, S. (**2015**). "The interaction of prosody and syntax in Samoan focus marking," Lingua **165**, 205–229.

Calhoun, S., Wollum, E., and Kruse Va'ai, E. (**2019**). "Prosodic prominence and focus: Expectation affects interpretation in Samoan and English," Lang. Speech. (published online)

Collins, P. C. (**1991**). *Cleft and pseudo-cleft constructions in English* (Routledge, London).

Collins, P. C. (**2006**). "*It*-clefts and *wh*-clefts: Prosody and pragmatics," J. Pragmat. **38**(10), 1706–1720.

Cooper, W. E., Eady, S. J., and Mueller, P. R. (**1985**). "Acoustical aspects of contrastive stress in question-answer contexts," J. Acoust. Soc. Am. **77**(6), 2142–2156.

Declerck, R. (**1988**). *Studies in Copular Sentences, Clefts, and Pseudo-Clefts* (Leuven University Press, Leuven, Belgium).

Destruel, E., Velleman, D., Onea, E., Bumford, D., Xue, J., and Beaver, D. (**2015**). "A cross-linguistic study of the non-at-issueness of exhaustive inferences," in *Experimental Perspectives on Presuppositions*, edited by F. Schwarz (Springer), pp. 135–156.

Eady, S. J., and Cooper, W. E. (**1986**). "Speech intonation and focus location in matched statements and questions," J. Acoust. Soc. Am. **80**(2), 402–415.

Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R., and Lotts, D. W. (**1986**). "Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments," Lang. Speech **29**(3), 233–251.

Gussenhoven, C. (**1983**). "Focus, mode and the nucleus," J. Linguist. **19**(2), 377–417.

Halliday, M. A. K. (**1967**). *Intonation and Grammar in British English* (Mouton, The Hague).

Hedberg, N. (**1990**). "Discourse pragmatics and cleft sentences in English," Ph.D. thesis, University of Minnesota, Minneapolis, Minnesota.

Kember, H., Choi, J., Yu, J., and Cutler, A. (**2019**). "The processing of linguistic prominence," Lang. Speech. (published online)

Krifka, M. (**2008**). "Basic notions of information structure," Acta Linguist. Hung. **55**, 243–276.

Kügler, F., and Calhoun, S. (**2020**). "Prosodic encoding of information structure: A typological perspective," in *The Oxford Handbook of Language Prosody*, edited by C. Gussenhoven and A. Chen (Oxford University Press, Oxford), pp. 454–467.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (**2017**). "lmerTest package: Tests in linear mixed effects models," J. Stat. Softw. **82**(13), 1–26.

Lambrecht, K. (**2001**). "A framework for the analysis of cleft constructions," Linguistics **39**(373), 463–516.

Lenth, R. (**2020**). "emmeans: Estimated marginal means, aka least-squares means," R package version 1.4.7, https://CRAN.R-project.org/package=emmeans (Last viewed 06/30/2020).

Luchkina, T. V. (**2016**). "Prosodic and structural variability in free word order language discourse," Ph.D. thesis, University of Illinois, Urbana-Champaign, Illinois.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (**2017**). "Balancing type I error and power in linear mixed models," J. Mem. Lang. **94**, 305–315.

Patil, U., Kentner, G., Gollrad, A., Kügler, F., Féry, C., and Vasishth, S. (**2008**). "Focus, word order and intonation in Hindi," J. South Asian Linguist. **1**(1), 55–72.

Pell, M. D. (**2001**). "Influence of emotion and focus location on prosody in matched statements and questions," J. Acoust. Soc. Am. **109**(4), 1668–1680.

R Core Team (**2020**). "R: A language and environment for statistical computing (version 4.0.0)," https://www.R-project.org/ (Last viewed 04/30/2020).

Rochemont, M. S. (**1986**). *Focus in Generative Grammar* (John Benjamins, Amsterdam).

Sánchez-Alvarado, C. (**2020**). "Syntactic and prosodic marking of subject focus in American English and Peninsular Spanish," in *Hispanic Linguistics: Current Issues and New Directions*, edited by A. Morales-Front, M. J. Ferreira, R. P. Leow, and C. Sanz (John Benjamins, Amsterdam), pp. 184–203.

Skopeteas, S., Féry, C., and Asatiani, R. (**2009**). "Word order and intonation in Georgian," Lingua **119**(1), 102–127.

Vainio, M., and Järvikivi, J. (**2006**). "Tonal features, intensity, and word order in the perception of prominence," J. Phon. **34**, 319–342.

Vainio, M., and Järvikivi, J. (**2007**). "Focus in production: Tonal shape, intensity and word order," J. Acoust. Soc. Am. **121**(2), EL55–EL61.

Wagner, M., and McAuliffe, M. (**2019**). "The effect of focus prominence on phrasing," J. Phon. **77**, 100930.

Xu, Y., and Xu, C. X. (**2005**). "Phonetic realization of focus in English declarative intonation," J. Phon. **33**(2), 159–197.