10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Prominence clash does not induce rhythmic adjustments in Italian

Francesco Burroni, Sam Tilsen

Department of Linguistics, Cornell University, Ithaca, USA

fb279@cornell.edu, tilsen@cornell.edu

Abstract

We present experimental evidence that, contrary to common belief, speakers of Italian do not adjust prominence to avoid clashes. Speakers of some languages (e.g. English, Italian) are believed to adjust prominence by shifting stress, or by deleting and/or inserting pitch accents. Although rhythmic adjustments may be produced in certain contexts (e.g. poetic verse, lexicalized phrases), we wondered whether naïve speakers produce them spontaneously. We used visual stimuli to elicit 3-word sequences with and without clash in two experiments with a total of 24 speakers of Italian. In both experiments we found no evidence for clash-induced adjustments. In Experiment 1, we observed a surprising increase in duration of the final syllable in the first word of a clashing pair accompanied by a very small decrease in intensity. No effects were observed for F0. In Experiment 2, we observed again a very small decrease in intensity of the initial syllable of the second word of a clashing pair. No effects were observed for duration or F0. These findings show that Italian speakers do not adjust prominence to avoid clash. Rather, clash induces localized increases in syllable duration. Since experimental evidence for rhythmic adjustments in English is also weak, we suggest that rhythmic adjustments may be a perceptual phenomenon, whose existence in production is constrained to specific contexts/lexical items.

Index Terms: rhythm, clash, stress, pitch accent, Italian

1. Introduction

An open question in research on speech prosody is whether speakers adjust prominence to make speech more rhythmic. The answer to this question remains elusive and hinges crucially on how rhythm is defined. It is, however, relatively uncontroversial that, in most languages, speech is not rhythmic "at the surface", that is, no units or events recur at regular temporal intervals (e.g. [1], [2], [3]).

In spite of this consensus, many theories of prosodic structure assume that speakers adjust prominence to make speech more rhythmic. One of the most prominent examples is avoidance of adjacent prominences (i.e. stress/pitch accent clash). Prominence clash plays an important role in various phonological analyses, where it is hypothesized to trigger stress shift, also known as the Rhythm Rule (e.g. [4]), and/or to cause pitch accent deletion with optional insertion of another pitch accent earlier in the phrase (e.g. [5]). Even though the shift and deletion accounts are somewhat different, it is clear that they harness the same intuition: prosodically prominent events should be spaced from one another, thus clashes are avoided.

This intuition has extended beyond the phonological literature to models of speech production. For instance, Levelt's influential model ([6]) stipulates an *ad hoc* lookahead mechanism to retrieve the metrical structure of an upcoming word to account for the (optional) application of the Rhythm Rule. Other scholars have proposed even stronger versions of rhythmic constraints. According to the Equal Spacing Constraint all stressed vowels tend to be attracted to periodically spaced temporal intervals ([7]). An identical regularizing mechanism for the spacing of prosodically prominent events has been proposed for speech perception ([8]). The search for rhythmic correlates of clashes in perception has marshalled EEG evidence to show that perceptually, the brain may be sensitive to clash ([9]).

However, experimental evidence that speakers actually adjust prominence to avoid clash is not compelling. Early work on English found evidence for decreased duration, F0 peak, and intensity of the final stressed syllable of the first word in a clashing pair (e.g. [10], [11]). However, a more recent study found these effects are manifested only in prepared speech and are quite subtle; neither duration nor intensity on the final stressed vowel of the first word in a clashing pair were affected by clash ([12]).

Another language in which rhythmic adjustments have been studied is Italian. An early phonetic study observed a decrease in duration of the final stressed syllable of the first word in a clashing pair ([13]). This study, however, was limited to only two speakers and only one of them exhibited the durational reduction. The study was also limited to prepared (read) speech.

Given that recent experimental work has cast doubt on the production of rhythmic adjustments in English, it is worthwhile to revisit the phenomenon in Italian.

2. Hypotheses and Predictions

Below we consider the predictions of three different models of rhythmic adjustment: (i) the Rhythm Rule (a.k.a. stress shift), (ii) pitch accent deletion and/or insertion, and (iii) prosodic break insertion.

2.1. The Rhythm Rule, a.k.a. Stress Shift

The Rhythm Rule (e.g. [4],[14]) hypothesizes that, in a clashing pair of words (henceforth w1 and w2), stress relocates from the final syllable of w1 ($w_1\sigma_f$) to the initial syllable of w1 ($w_1\sigma_i$). For example, *fourTEEN FLOORS* would be adjusted to *FOURteen FLOORS (cf. floor number fourTEEN)* and Italian *ciTTÀ* SPOrca would be adjusted to *CItà SPOrca*. This stress shift account predicts that the acoustic correlates of stress (i.e. duration, intensity, and/or F0) should be diminished for σ_f of w1 and enhanced for σ_i of w1.

10.21437/SpeechProsody.2020-117

2.2. Pitch Accent Deletion and/or Insertion

Pitch accent deletion and/or insertion accounts hypothesize that the pitch accent associated with σ_f of w1 is deleted in clash and a pitch-accent may be optionally inserted on σ_i of w1. The prediction of this account is a decrease in F0 of σ_f of w1, and, optionally an increase in F0 of σ_i of w1.

2.3. Prosodic break insertion

Another clash avoidance mechanism that has been mentioned in the phonological literature is the insertion of a prosodic boundary or a pause. This pause or boundary would have the effect of spacing adjacent prominences further away from each other.

The specific predictions of this account depend on the nature of the hypothesized pause or prosodic boundary. If the prosodic break is really a pause it should be detected as silence in the acoustic signal. However, if a prosodic boundary is inserted, we could expect that the σ_f of w1 is lengthened and F0 is reduced. Moreover, σ_i of w2 after the boundary may be lengthened and exhibit higher F0 due to pitch reset. These are the most common acoustic correlates of prosodic boundaries in English (e.g. [15]), however, it is not known from experimental work whether Italian prosodic boundaries have acoustic correlates comparable to the English ones.

3. Methods

3.1. Experiment 1

16 native speakers of Italian (8M, 8F) participated in Experiment 1. The experimental design was inspired by an earlier investigation of the Rhythm Rule in English ([12]). Participants sat in a sound-attenuated room in front of a monitor. On each trial, three visual stimuli were presented, corresponding to a numeral (w0), a noun (w1), and a color adjective (w2), (see Table 1 below). Participants were instructed to produce the three-word phrase corresponding to the stimuli.

Table 1: *Stimuli for Experiment 1.*

w0	w1	w2	
DU.e 'two'	**ca.FFÈ** 'coffee'	bor.DO 'bordeaux'	no clash
NO.ve 'nine'	**ci.TTÀ** 'city'	ma.RRO.ni 'brown'	
MI.lle 'thousand'	**co.li.BRÌ** 'hummingbird'	*NE.ri* 'black'	*clash*
		VER.di 'green'	

Participants produced the 36 unique three-word combinations of Table 1 in 10 blocks separated by short breaks. One participant completed only 8 blocks due to time constraints on experiment duration after an equipment failure. In odd blocks participants were instructed to mentally rehearse the utterance before producing it, in even blocks participants began production immediately. Since this manipulation had no effect on the analyses presented here, it is not further discussed. Participants were also instructed to try to maintain a constant speech rate, to speak clearly and informally - as if they were talking to a friend-, and to try producing target utterances with the intonation of a declarative utterance.

The target word of experiment 1 was w1 (bolded in Table 1), while w2 was used to manipulate the clash (italics in Table 1). We analyzed acoustic measurements for σ_i and σ_f of w1, as well as the vowels contained in these syllables: V_i and V_f.

3.2. Experiment 2

8 native speakers of Italian (4M,4F), who had participated in Experiment 1, participated in Experiment 2 approximately 6 months later. The experimental setting and task were the same as in Experiment 1. However, the visual cues differed to represent the targets in Table 2.

Table 2: *Stimuli for Experiment 2.*

w0	w1	w2
DU.e 'two'	CA.li.bri 'calibers'	bor.DO 'bordeaux'
NO.ve 'nine'	co.LU.bri 'adders'	ma.RRO.ni 'brown'
MI.lle 'thousand'	*co.li.BRÌ* 'hummingbird(s)'	***NE.ri*** 'black'
		VER.di 'green'

The target of experiment the target was w2 (bolded in Table 2), while both w1 and w2 (italics in Table 2) were used to manipulate distance between the final prominence of w1 and the initial prominence of w2 (a distance of 0 corresponds to clash). Since our main concern are cases of clash, we analyzed acoustic measurements for σ_i of w2 with initial stress only (*NE.ri*, *VER.di*) after all combinations of w1, as forms with non-initial stress cannot result in a clash.

W2 forms that cannot result in clash were used to replicate the findings of Experiment 1 by comparing w1 *co.li.BRÌ* in no-clash vs clash. The replication holds; however, it is not presented here for reasons of space and because this adds nothing to the picture emerging from the results of Experiment 1.

3.3. Data processing and analysis

Audio was recorded at 22.05 kHz using a head-mounted microphone. To obtain durations and facilitate other acoustic measurements, HMMs for forced alignment were trained in Kaldi ([16]), using 18 manually segmented trials from each participant All trials were subsequently aligned. The extracted acoustic measurement of interest are as follows:

- Duration of σ_i and of σ_f of w1 (Experiment 1) and σ_i of w2 (Experiment 2).
- RMS intensity ratio of the final vowel over the initial vowel V_f/V_i of w1 (Experiment 1); and raw RMS intensity of V_i of w2 (Experiment 2).
- Median F0 of V_i and V_f of w1 (Experiment 1) and of the initial V_i of w2 (Experiment 2).

Duration, RMS intensity, and F0 were chosen as the most commonly measured correlates of word- and phrasal level prominence in Italian and other languages (e.g. [18],[19]). Duration was calculated using segmental boundaries of the forced alignments. RMS intensity was calculated as the root mean squared value of the signal over 25 ms windows with a 5 ms overlap between each window. F0 values were calculated using a MATLAB implementation of Talkin's robust algorithm for pitch tracking (RAPT [20]) contained in Voicebox ([21]).

Before statistical analysis, mean and standard deviation (std) of each acoustic measurement were calculated separately for each speaker and all data points exceeding 2 std (duration) or 3 std (RMS intensity) from the mean were excluded from subsequent analyses. For F0 processing, in all trials all points exceeding 2 std from the mean or deviating ±10 Hz from the preceding sample were removed. F0 was subsequently linearly interpolated and smoothed using a median filter. The counts of collected and excluded data points are recapitulated in Tables 3 and 4.

Table 3: *Data collected and excluded in Experiment 1*

Measurement	**Experiment 1**	
	Total	Excluded
Duration σ_i	5688	221 (3.8%)
Duration σ_f	5688	278 (4.9%)
RMS Intensity V_f/V_i	5688	643 (11%)
F0 V_i	5688	404 (7%)
F0 V_f	5688	449 (7.8%)

Table 4*: Data collected and excluded in Experiment 2*

Measurement	**Experiment 2**	
	Total	Excluded
Duration σ_f	1432	43 (3%)
RMS V_i	1432	3 (.2%)
F0 V_i	1432	2 (.1%)

Linear mixed effect regressions were fitted to the data using the models (1) and (2) below, for Experiment 1 and 2, respectively. Speaker and word were used as random effects, and clash or interstress distance (with three levels 0 syllable distance (clash), 1, and 2) were used as the fixed effects. The models were compared against intercept only models.

$$\textit{variable of interest} \sim \textit{clash} + (1|\textit{subject}) + (1|\textit{word}) \quad (1)$$

$$\textit{variable of interest} \sim \textit{stress dist} + (1|\textit{subject}) + (1|\textit{word}) \quad (2)$$

Post-hoc testing was conducted using one-way ANOVA and Tukey HSD.

4. Results

4.1. Experiment 1

Duration. Clash was found to have a main effect on the duration of σ_i and of w1 (χ^2 =33.48, p < .001). However, the magnitude of the effect size was very small, estimated at 2 ms, much below the just noticeable difference threshold of 10 ms for auditory stimuli shorter than 240 ms ([22]), see Figure 1 left.

Clash was also found to have a main effect on the duration of σ_f of w1 (χ^2 = 693.11, p < .0001). However, contrary to the stress retraction hypothesis, clash was associated with an *increase* in the duration of σ_f, with effect size estimated at 16 ms, see Figure 1 right.

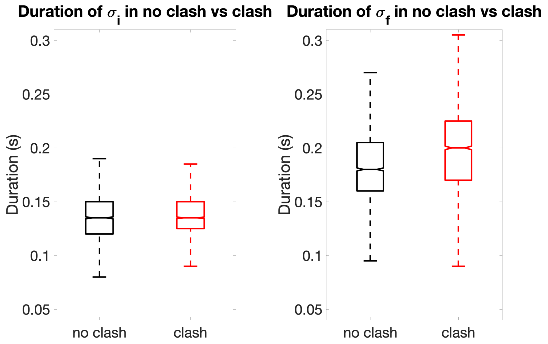

Figure 1: *Duration of σ_i and σ_f in no clash vs clash*

RMS Intensity. Clash had a main effect on V_f / V_i intensity ratio (χ^2 = 18.74, p < .0001), with an effect size estimated at -0.09. That is, clash caused a very slight decrease in the intensity of σ_f relative to σ_i.

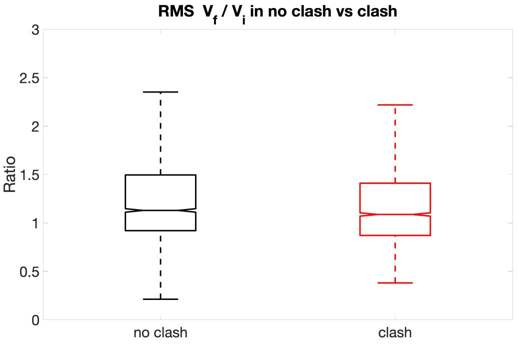

Figure 2: *RMS intensity of V_f/V_i.*

F0. Clash has no main effect on the median F0 of V_i (χ^2 = 1.69, p = .19). Clash has no main effect on the median F0 of V_f either (χ^2 = 2.63, p = .10).

The overall effects of clash in the investigated acoustic dimensions are summarized in Table 5.

Table 5: *Effects of clash on w1 of a clashing pair.*

w1	**Duration**	**w1**	**RMS Intensity**	**w1**	**F0**
σ_i	↑ +2ms	V_f /V_i	↓ -0.09	V_i	-
σ_f	↑ +16ms			V_f	-

4.2. Experiment 2

Duration. The presence of a clash had no main effect on the duration of σ_i of w2 (χ^2 = 1.43, p = .23).

RMS Intensity. The presence of a clash had a marginally significant main effect on RMS intensity of σ_i in w2 (χ^2 = 3.96, p = .04). The effect size is estimated at -0.0045.

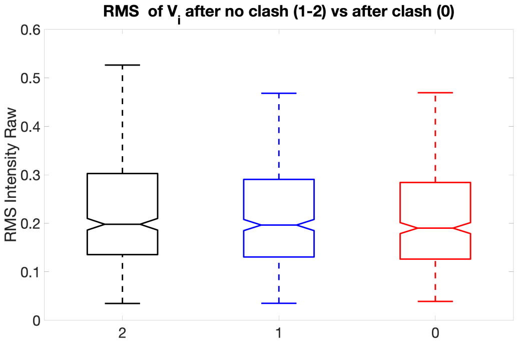

Figure 3: *RMS intensity as a function of interstress distance: no clash (1 -2 σ distance) vs clash (0 σ distance).*

F0. The presence of a clash had no main effect on the median F0 σ_i of w2 in a clashing pair ($\chi^2 = 0.29$, p = .58).

The overall effects of clash in the investigated acoustic dimensions are summarized in Table 6.

Table 6: *Effects of clash on w2 of a clashing pair.*

w2	Duration	w2	RMS Intensity	w2	F0
σ_i	-	V_i	↓ -0.0045	V_i	-

5. Discussion

The experimental results show that in Italian the main effect of clash on w1 of a clashing pair is, contrary to expectation, a significant lengthening of σ_f, while the lengthening on σ_i is negligible. The lengthening of σ_f was accompanied by a small decrease in intensity. No effect was observed on F0. The increased lengthening of σ_f is hard to explain if clashes indeed are repaired with a process of stress retraction. Similarly, the lack of effects on F0 of σ_i and σ_f in w1 is not consistent with the predictions of an accent deletion account, in which a pitch accent is deleted on σ_f and another is optionally inserted on σ_i.

The lengthening of σ_f in w1 is compatible with the insertion of a prosodic boundary. To verify the plausibility of this hypothesis, in experiment 2, we tested whether clash may also affect σ_i of w2 in a clashing pair. We did not observe any durational of F0 effects of clash on w2, effects that may be expected under a boundary insertion account. The only effect that we observed was a small decrease in RMS intensity of σ_i of w2.

Given that the largest effect of clash observed across experiments was *lengthening* of a syllable involved in clash, we interpret this result as evidence against all three mechanisms of rhythmic adjustment, which predicts shortening of this syllable.

There were also small effects of clash on the intensity of the clashing syllables, but these effects were not fully consistent with the predictions of any account. Specifically, although the boundary insertion and stress shift accounts do predict clash-induced decrease of intensity in σ_f of w1, these accounts also predict increased intensity in σ_i of w2, contrary to findings. Thus, the intensity patterns provide only equivocal evidence for rhythmic adjustment models.

In sum, the idea that prominence clashes trigger rhythmic readjustments in Italian is problematic. Evidence that prominence profiles are altered, or boundaries inserted to avoid clash is lacking. We suggest that the experimental evidence should be taken at face value and that the main correlate of prosodic prominence clash in Italian is a localized delay in the production of σ_f of w1 of a clashing pair. This picture is compatible with recent work on clashes in English, where it has been shown ([12]) that clash correlates with increased duration of σ_f of w1 of a clashing pair in relatively unprepared speech and that boosts on σ_i of w1, if present at all, are very weak.

The mechanisms responsible for the lengthening effect are of substantial interest. One possibility is that speakers prolong the period of time that gestures in σ_f of w1 are active, because the clash causes them to attend more closely to external sensory feedback of their own speech ([24]). Alternatively, it is possible that planned prominences may interact, and this interaction may result in an activation boost on the first prominence of a clash, which in turn influences gestures that are associated with it. A proposal along these lines would be compatible with hypotheses previously formulated in the literature (e.g. [12]).

To conclude, we emphasize that our results contrast sharply with a wide body of literature reporting the impression of prominence shift for Italian, English, and other languages. The mismatch between production results and perceptions begs for an explanation. As it has already been pointed out in the perception literature, prominence shift perception seems to be a highly context-dependent perceptual illusion ([8]). This perceptual illusion need not be the only source for the report of prominence shifts by trained linguists and naïve listeners alike. It is also possible that prominence shift exists in highly specialized production contexts. For instance, highly lexicalized expression may have different prominence profiles than the citation forms of their individual parts. For example, It. *meTÀ* ‘half’ may be highly lexicalized as *meta* (with no prominence) in forms like *meta STRAda* ‘halfway’. Alternatively, it is possible that that prominence adjustment may originate in poetic verse and extend to poetic-like rhetorical speech (e.g., [25]), where prominence adjustments are often mandatory. Our work, however, casts doubt on the idea that rhythmic adjustments are naturally occurring phenomena in spontaneous speech.

6. Conclusion

In this paper we have presented experimental evidence from Italian showing that the main correlate of prominence clash in this language is a lengthening of the final syllable of w1 in a clashing pair. This is compatible with recent work on English. It is thus possible that the lengthening effects may be a cross-linguistic correlate of prominence clash. This is a question that is left open for future research together with the mechanisms responsible for this effect.

Importantly, the observed effects are at odds with previous rhythmic accounts of prominence clashes. Contrary to the claims found in a wide body of literature on Italian and English, clashes do not induce obvious rhythmic readjustments in speakers’ production. If rhythmic constraints are indeed enforced on speech production, it is not clear that prominence clashes can be marshalled as evidence in favor of their existence, at least in Italian and English.

7. Acknowledgements

We are grateful to Abigail Cohn and Draga Zec, as well as the member of the Cornell Phonetics laboratory and four anonymous reviewers for precious feedback on earlier versions of this work.

8. References

[1] A. Turk and S. Shattuck-Hufnagel. “What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapic, and Goswami and Leong”. *Laboratory Phonology*, vol. 4, no. 1, pp. 93–118, 2013.

[2] F. Cummins. “Rhythm and speech”. In M. Redford (ed.), *The Handbook of Speech Production*, pp. 158–177. John Wiley & Sons, 2015.

[3] F. Nolan and J. Hae-Sung. “Speech rhythm: a metaphor .” *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1658 (2014): 20130396.

[4] M. Liberman and A. Prince. "On stress and linguistic rhythm." *Linguistic inquiry* 8.2 (1977): 249-336.

[5] S. Shattuck-Hufnagel, M. Ostendorf, and K. Ross. "Stress shift and early pitch accent placement in lexical items in American English." *Journal of Phonetics* 22.4 (1994): 357-388.

[6] W.J.M. Levelt. *Speaking: From Intention to Articulation*. MIT press, 1993.

[7] H. Quené and R. F. Port. "Rhythmical factors in stress shift." *38th meeting of the Chicago Linguistic Society: The main session*. 2002.

[8] J.M Tomlinson, Q. Liu, and J. E. Fox Tree. "The perceptual nature of stress shifts." *Language, Cognition and Neuroscience* vol. 29, n0 9, pp. 1046-1058, 2014.

[9] K. Bohn, et al. “The influence of rhythmic (ir)regularities on speech processing: evidence from an ERP study on German phrases.” Neuropsychologia, vol. 51, no. 4, pp. 760-771, 2013.

[10] I. Vogel, H. T. Bunnell, and S. Hoskins, “The phonology and phonetics of the Rhythm Rule,” in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, B. Connell and A. Arvaniti, Eds. Cambridge: Cambridge University Press, 1995, pp. 111–127.

[11] E. Grabe and P. Warren, “Stress shift: do speakers do it or do listeners hear it?” in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, B. Connell and A. Arvaniti, Eds. Cambridge: Cambridge University Press, 1995, pp. 95–110.

[12] S. Tilsen. "Utterance preparation and Stress Clash: Planning prosodic alternations." *Speech production and perception: Planning and dynamics. Peter Lang Verlag*. (2012).

[13] E. Farnetani and S. Kori. "Interaction of syntactic structure and rhythmical constraints on the realization of word prosody." *Quaderni del Centro di Studio per le Ricerche di Fonetica*, vol. 2, pp. 288-318, 1983.

[14] M. Nespor, and I. Vogel. “Clash avoidance in Italian.” *Linguistic Inquiry*, vol. 10, no. 3, pp. 467-482, 1979.

[15] A. M. Brugos. *The interaction of pitch and timing in the perception of prosodic grouping*. Ph.D. Dissertation, Boston University, 2015.

[16] D. Povey et al. "The Kaldi speech recognition toolkit." *IEEE 2011 workshop on automatic speech recognition and understanding*. No. CONF. IEEE Signal Processing Society, 2011.

[17] P. Boersma and D. Weenink. Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 14 March 2018 from http://www.praat.org/

[18] A. Eriksson et al. “The acoustics of lexical stress in Italian as a function of stress level and speaking style.” *INTERSPEECH 2016, San Francisco, USA, September 8–12, 2016*. International Speech Communication Association, 2016.

[19] V. J. van Heuven, “Acoustic Correlates and Perceptual Cues of Word and Sentence Stress: Towards a Cross-Linguistic Perspective,” in *The Study of Word Stress and Accent: Theories, Methods and Data*, Cambridge: Cambridge University Press, 2018, pp. 15–59.

[20] D. Talkin. “A robust algorithm for pitch tracking (RAPT).” In *Speech Coding and Synthesis*, pp. 497-518, 1995.

[21] M. Brookes. “VOICEBOX: A speech processing toolbox for MATLAB. 2006.” *URL http://www.ee.ic.ac.uk/... hp/staff/dmb/voicebox/voicebox. Html.*

[22] A. Friberg, and J Sundberg. “Perception of just-noticeable time displacement of a tone presented in a metrical sequence at different tempos.” *The Journal of The Acoustical Society of America*, vol. 94, no. 3, pp. 1859-1859, 1993.

[23] M Nespor and I. Vogel. “On clashes and lapses.” *Phonology*, vol. 6, no. 1, pp. 69-116, 1989.

[24] S. Tilsen. “Space and time in models of speech rhythm”. *Annals of the New York Academy of Sciences*, vol. *1453*, no. 1, pp.47-66, 2019.

[25] C. Gussenhovel. “Stress shift in Dutch as a rhetorical device” *Linguistics*, vol. 21, no. 4, pp. 603-620, 1983.