Interspeech 2024
1-5 September 2024, Kos, Greece

Do we EXPECT TO find phonetic traces for syntactic traces?

Jonathan Him Nok Lee[1], *Mark Liberman*[1], *Martin Salzmann*[2]

[1]University of Pennsylvania
[2]University of Potsdam
jonhnlee@sas.upenn.edu, myl@cis.upenn.edu, salzmann2@uni-potsdam.de

Abstract

In syntactic theories, the existence of silent/inaudible syntactic elements, such as movement traces, has been hypothesized to impact phonetic outcomes. One classic example involves the contraction of "want to" into "wanna," where intervening syntactic traces are thought to prevent such contraction. This study extends this inquiry to "expect to," a similar construction but without established morpho-phonological contractions, and investigates whether there are any phonetic effects of the presence of traces. Drawing on findings from *wanna*-contraction, syntactic theories predict that intervening traces will hinder phonetic reduction in "expect to." To investigate, we randomly sampled 300 utterances containing "expect to" from a corpus of NPR podcasts. Multivariate linear regression shows no relationship between the presence of intervening syntactic traces and the duration of any phones in "expect to." However, multinomial logistic regression showed that the lenition of "expect to" ("to" /tu:/ → [tʌ] or [tɪ]) was significantly more likely to occur across intervening syntactic traces. Our findings contradict the predictions posited by syntactic theories, suggesting a more intricate interplay between syntax and phonetics.

Index Terms: expect to, syntactic trace, phonetic reduction, duration, lenition

1. Introduction

Modern syntactic theories have proposed the existence of silent or inaudible syntactic elements, including traces of syntactic movements and PRO (a pronoun without phonological realization; not a syntactic trace). These elements are motivated by syntax-internal considerations. Exploring the (indirect) effects of these silent elements in pronounced structures can offer external evidence supporting their presence.

The English *wanna*-contraction phenomenon serves as a classic example supporting the argument that the effects of traces from syntactic movements can intervene in the process of contraction. *Wanna*-contraction, initially explored by Lakoff [1], is illustrated through the following minimal pair:

(1) Teddy is the man I *want to* succeed.
(2) Teddy is the man I *wanna* succeed.

In (1), two interpretations exist: (a) I want Teddy to succeed and (b) I want to succeed Teddy. In contrast, (2) is exclusively interpreted as (b) I want to succeed Teddy. Syntactically, a movement trace of "Teddy" occurs between "want" and "to" in interpretation (a). The syntactic trace in interpretation (b) appears after "succeed". It is argued that the intervening trace between "want" and "to" in interpretation (a) prevents the *wanna*-contraction, while the silent PRO in interpretation (b) does not prevent the contraction. Consequently, only interpretation (b) is viable in (2), while both interpretations are possible in (1), as the trace of "Teddy" can be recovered either between "want" and "to" or after "succeed" [1], [2].

Interpretation (a)
Teddy$_i$ is the man I *want* ___$_i$ *to* succeed.
(= I want Teddy to succeed.)

Interpretation (b)
Teddy$_i$ is the man I *want* PRO *to* succeed ___$_i$.
(= I want to succeed Teddy.)

Despite the consistent grammatical judgment and semantic interpretation across different English varieties [2], Liberman [3] identified exceptions in a large corpus of 105,817 transcribed NPR podcasts, where *wanna*-contraction occurred across a trace in spontaneous speech. He concluded that the morphosyntactic generalization for *wanna*-contraction remains statistically true but not categorically in phonetic terms. He suggested that there can be two different sources of *wanna*-contraction: a set of regular (optional, gradient) phonetic lenition phenomena; and a lexicalized phonological representation of the fully-lenited form [4]. The exceptions of *wanna*-contraction are potentially phonetic lenition rather than a morphosyntactic substitution.

Crucially, *wanna*-contraction and other forms of contraction (such as finite auxiliary reduction) exhibit specific morpho-phonologically determined reduction outcomes. For instance, "want to" can only be contracted as "wanna," and "have" can be reduced to "'ve." However, it remains unexplored whether typical phonetic reduction, characterized by no specific reduction outcomes, is also influenced by similar restrictions imposed by syntactic environments. Given the observed constraints on contraction due to intervening syntactic traces, our study aims to explore the connections between phonetic reductions devoid of fixed/lexicalized contraction forms and intervening syntactic traces.

In the case of *wanna*-contraction, only the intervening syntactic trace in interpretation (a), but not PRO in interpretation (b), can constrain contraction possibilities. Unlike "want," it is not generally the case that verbs allow both constructions. However, similar syntactic environments are present in the construction of "expect to", which is compatible with both exceptional case-marking (ECM)-parse like "want to" and Control-parse (PRO). This compatibility is illustrated respectively by examples (3) and (4) extracted from our analyzed data transcripts. In the word sequences "expect to,"

10.21437/Interspeech.2024-2190

there can be an intervening syntactic trace in (3) caused by the extraction of "who" (i.e., relativization) or a silent PRO in (4) between "expect" and "to." There are no established morpho-phonologically contraction or reduction outcomes for "expect to."

(3) You don't necessarily like the man$_i$ who you *expect* ___$_i$ *to* be trying to implement them.

(4) I want you to just write what you *expect* PRO *to* gain out of this kind of mentorship.

Drawing on findings from *wanna*-contraction, syntactic theories predict that intervening traces will impede phonetic reduction in "expect to." The literature of form variations has suggested that duration and reduction frequency are the main differences between surface forms and different lemmas [5]. We hypothesize that phonetic reduction, including shortened duration and lenition, is less likely to occur in "expect to" across syntactic traces (as in (3)) compared to situations without intervening traces (e.g., PRO (4)).

2. Method

2.1. Data collection and procedure

To investigate potential phonetic effects related to syntactic traces, we randomly sampled 300 utterances containing the word sequences "expect to" from a large corpus of 105,817 transcribed NPR podcasts. This corpus was previously analyzed by Liberman [3] in his study of *wanna*-contraction. We first determined whether each utterance had an intervening syntactic trace between "expect" and "to" or simply a non-trace PRO. We also checked that all utterances with intervening traces had the same syntactic structure and operation (i.e., A'-movement for relativization). After excluding reduplicated utterances, mistranscribed utterances, and those with poor sound quality, we incorporated a total of 256 utterances into our analysis. We investigated two possible phonetic effects brought by syntactic traces: shortened duration and lenition.

2.1.1. Duration

We performed forced alignment on our data using Penn Phonetics Lab Forced Aligner (P2FA) [6] to time-align each phone of the words based on transcription to the audio recordings. Subsequently, with the aid of waveforms and spectrograms in Praat [7], a phonetician manually checked and adjusted the time alignments of the phones of the word sequences "expect to" in all analyzed utterances and measured the duration of each phone of "expect to" using ProsodyPro [8]. Figure 1 illustrates an example of our segmentation of "expect to." From top to bottom, the first panel displays the waveforms of the audio recording, and the second presents the spectrogram. The third panel (i.e., the first annotated tier) depicts the phones of the transcribed words, and the fourth panel shows the transcribed words. The words "expect to" typically involve the following phone sequences: /ɪkspekt tu:/. The first /k/ was delimited by the end of the vowel formants of the preceding /ɪ/ and the beginning of the fricative noise of the subsequent /s/. /p/ was delimited by the end of the fricative noise of the preceding /s/ and the beginning of the vowel formants of the vowel /e/. The codas -/kt/ in "expect" were mostly unreleased in our data, annotated by one coda label -/kt/ due to the absence of clear signals of the boundary between /k/ and /t/. The initial /t/- in "to" was consistently aspirated in our data. Its boundary was delimited by the start of the aspiration and the beginning of the voicing of the subsequent vowel. The end of the vowel of "to" was defined by the end of vowel formants.

2.1.2. Lenition

In addition to duration, we explored the lenition phenomena of the vowel of "to." While the fully pronounced form of "to" is /tu:/, reduction of the long vowel /u:/ into /ʌ/ and /ɪ/ is common in spontaneous speech. P2FA classified the vowels of "to" based on the pre-trained data in their Gaussian Mixture Models-based Hidden Markov Model. To enhance accuracy, a phonetician cross-verified and adjusted P2FA's classifications using perception judgments and the information from waveforms and spectrograms.

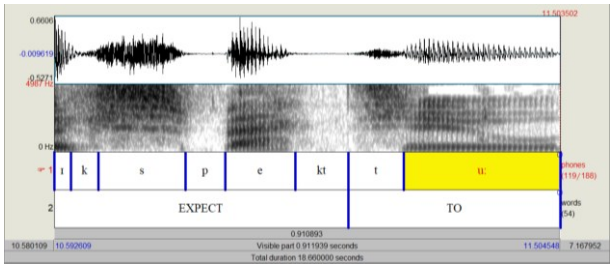

Figure 1: *Example of segmenting "expect to".*

2.2. Data analysis

In terms of phone duration, it appeared that the structure of our data did not involve nested random factors as utterances in our analyzed data were produced almost exclusively by distinct speakers in the NPR podcasts. Multivariate linear regression was conducted on our data in R [9]. In our regression model, the duration of each phone served as the dependent variable predicted by the binary variable of syntactic trace (presence vs. absence). Our multivariate analyses allowed the coefficients from each model to covary [10], meaning that covariance was considered when determining whether the predictor, syntactic trace, jointly contributed to the model of each phone in "expect to." Standardized β was calculated as a measure of effect size [11], [12].

For the analyses of lenition in "to," as there were multiple unordered lenition possibilities, we performed multinomial logistic regression on the lenition data using the R package *nnet* [13]. Multinomial logistic regression is a versatile multivariate approach that does not presume normality, linearity, or homoscedasticity [14]. The dependent variables were various forms of the vowel in "to." We configured the log-likelihood of producing /u:/ (fully pronounced form) as the reference level using dummy coding in our model. The predictor was the binary variable of syntactic trace (presence vs. absence). Additionally, we included the duration of the consonant and vowel in "to" and their interactions as covariates in the models because their durations likely covary with different forms of the vowel in "to," whether fully pronounced or reduced forms. Relative risk ratio was calculated by exponentiating the coefficients of the models. Relative risk ratio in the current study was defined as the ratio of the probability of producing a particular lenition form versus the probability of producing the fully pronounced form /u:/. Analyses of vowel formants were not considered in this study because vowels were sometimes undetectable due to deletion (see section 3.2).

3. Results

Of the 256 analyzed utterances, 65 (25.4%) involved syntactic traces between "expect" and "to," while 191 (74.6%) did not. Interestingly, we found two instances of speakers inserting a filled pause "uh" between "expect" and "to," and they were exclusive to the utterances involving an intervening syntactic trace between "expect" and "to," despite utterances having intervening traces being the minority in our dataset. We excluded these two utterances from our subsequent statistical analyses as the disfluent utterances are not comparable to other analyzed fluent utterances.

3.1. Duration

Figure 2 compares the duration of each phone in "expect to" with intervening syntactic traces (Trace = Yes) and without (Trace = No). Table 1 reports the results of our multivariate linear regression. The presence of syntactic traces had no significant effects on the duration of any of the phones in "expect to." The standardized β of the model was .04, indicating negligible effect size of the model [11].

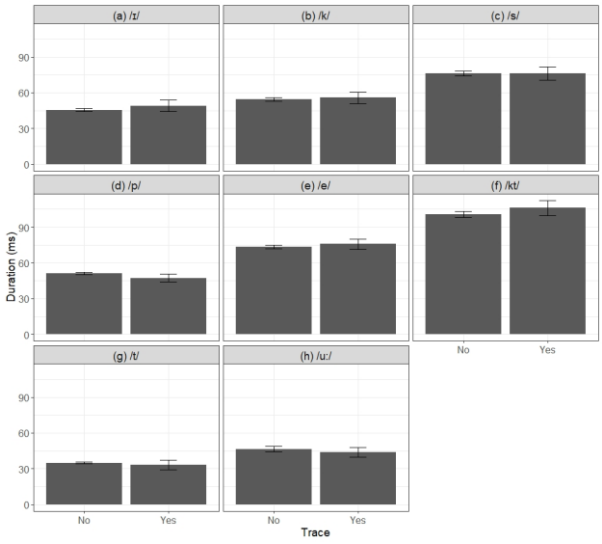

Figure 2: *The duration of each phone in "expect to" with (Yes) and without (No) syntactic traces.*

Table 1: *Results of multivariate linear regression*.

Dependent variable	*B*	*SE*	*t*	*p*
/ɪ/	1.70	2.90	.59	.56
/k/	2.36	3.30	.72	.48
/s/	-1.04	4.27	-.24	.81
/p/	-2.61	2.20	-1.19	.24
/e/	-1.40	3.11	-.45	.65
/kt/	2.67	5.31	.50	.62
/t/	-2.33	2.12	-1.10	.28
/u:/	-.27	4.62	-.06	.95

3.2. Lenition

Three types of lenitions in the vowel /u:/ of "to" were attested: /ʌ/, /ɪ/, and deletion of the vowel. Figure 3 illustrates an example of deleting the vowel in "to". In the highlighted region between the noise of the aspiration of /t/- and the aspiration of /p/- of the subsequent word, it was spectrally clear that no voicing was detected, let alone vowel formants. Reading from the waveform and spectrogram suggests that the speaker only pronounced an aspirated /t/ without a vowel for the word "to."

Table 2 provides an overview of the distribution of vowel forms in "to." The predominant form is /ʌ/ (51.56%), observed with or without an intervening trace (49.23% and 52.36%, respectively). The fully pronounced form /u:/ was less frequent than both /ʌ/ and /ɪ/, comprising a total of 14.06%. The least common occurrence was the deletion of the vowel (3.13%).

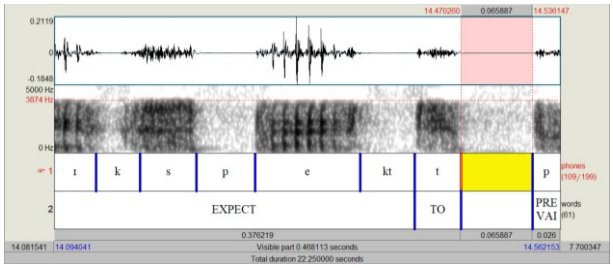

Figure 3: *Example of vowel deletion in "to".*

Table 2: *Distribution of the vowel forms in "to".*

Form	With trace (*n* = 65)	Without trace (*n* = 191)	Total (*n* = 256)
/u:/	7 (10.78%)	29 (15.18%)	36 (14.06%)
/ʌ/	32 (49.23%)	100 (52.36%)	132 (51.56%)
/ɪ/	24 (36.92%)	56 (29.32%)	80 (31.25%)
Deletion	2 (3.08%)	6 (3.14%)	8 (3.13%)

Owing to the limited number of observations for vowel deletion (< 10), it was excluded from our multinomial logistic regression model because of insufficient sample size [14]. Table 3 reports the results of our multinomial logistic regression. The likelihood of producing /u:/ (full pronounced form) served as the reference level in the model. Our primary predictor, the presence of intervening syntactic traces (With trace), was significantly associated with reducing vowels to both /ʌ/ (p = .03) and /ɪ/ (p < .001). The relative risk ratios of syntactic traces in both levels were greater than 1, indicating that having an intervening trace was a risk factor for lenition. For instance, the relative risk ratio 1.40 of trace for /ʌ/ implies that, with an intervening trace, the odds (relative probabilities) of reducing vowels to /ʌ/ increased by a factor of 1.40 compared to the reference level /u:/. In other words, the odds of producing /ʌ/ raised by 40% when there was an intervening trace compared to when there was not. Similarly, the relative risk ratio 1.86 for /ɪ/ suggests that, with an intervening trace, the odds of producing /ɪ/ instead of /u:/ increased by a factor of 1.86. In this case, the odds of reducing vowels to /ɪ/ increased by 86% when there was a trace compared to when there was no trace.

The effects of other covariates, such as the duration of /t/- and vowels, on lenition forms were significant in both /ʌ/ and /ɪ/ (all p < .001), primarily due to the inherently reduced nature of lenition forms, characterized by shorter duration. Despite their statistical significance, these variables showed very small effect sizes (unstandardized *B* between -.05 and -.09; relative risk ratio between .92 and .95), consistent with the non-significant results in the multivariate linear regression presented in Table 1. Since effect size is independent of sample size [15], it suggests that our two statistical models were not affected by

Table 2: *Results of multinomial logistic regression*

Form	Predictor	*B*	*SE*	*z*	*p*	Relative risk ratio
/ʌ/	With trace	.34	.16	2.12	.03*	1.40
	Duration of /t/-	-.05	.02	-3.66	<.001***	.95
	Duration of vowel	-.09	.01	-8.06	<.001***	.92
	Duration of /t/-* Duration of vowel	.00	.00	1.80	.07	1.00
/ɪ/	With trace	.62	.16	3.98	<.001***	1.86
	Duration of /t/-	-.06	.02	-3.33	<.001***	.94
	Duration of vowel	-.07	.01	-6.08	<.001***	.93
	Duration of /t/-* Duration of vowel	.00	.00	.82	.41	1.00

* $p < .05$, *** $p < .001$.

the lack of statistical power. The interaction effects between the duration of /t/- and vowels were, however, non-significant.

4. Discussion

Our results reveal no relationships between intervening syntactic traces and the duration of any phones in "expect to." However, there are significant associations between the presence of intervening traces and the vowel lenition in "to." Intervening traces are significant risk factors for vowel lenition in "to". This suggests that phonetic reduction is more likely to occur across syntactic traces than without intervening traces (e.g., across PRO). These findings challenge the predictions made by syntactic theories, which posited the opposite. Drawing on the findings of *wanna*-contraction, syntactic theories predicted that traces would restrict contraction possibilities.

We speculate that there might be moderating effects of the typical unreleased -/kt/ coda of "expect," which might influence how speakers articulate "to" in the word sequence "expect to." Moreover, with respect to the lenition of "to," there may be a differential binding strength forward (to the following verb) rather than backward (to "expect"). Consequently, the results align reciprocally with the presence of syntactic traces. Future studies should replicate the lenition possibilities in other comparable exceptional case-marking (ECM) verbs to determine if there are consistent patterns. Moreover, recent reviews have highlighted significant mismatch in syntactic and prosodic structures [16], [17]. Studies in production planning have also emphasized the impact of syntactic and prosodic boundaries, as well as word frequency, on production variability [18], [19], [20]. Additional research could analyze how the position of "to" within prosodic and syntactic phrases, whether final or non-final positions, along with the frequency of the subsequent word, influences the likelihood of lenition.

Even though our findings contradict the predictions of syntactic theories, they are still consistent with the hypothesis that silent or inaudible syntactic elements can impact phonetic outcomes, suggesting a more intricate interplay between syntax and phonetics. It is still strikingly intriguing to see how abstract syntactic traces can affect the possibilities of non-morpho-phonologically determined phonetic reduction in real time (as in naturalistic spontaneous speech). Alternatively, phonetic reduction in our case can support the existence of syntactic traces. Meanwhile, in the 1980s, a few alternative proposals sought to explain *wanna*-contraction without resorting to trace theory [21]. While these proposals were dismissed by syntacticians in the 1990s, it is worth noting that modern syntactic theories keep evolving, and their conceptualization of trace may differ in the future [22], [23], [24]. Given that the results of the present study challenge the predictions of current syntactic theories, future syntacticians may revisit the phonetic evidence for *wanna*-contraction and "expect to" and refine their theories accordingly.

We recognize that the generalizability of the current study may be limited by the small sample size. Nevertheless, we believe that our study serves as a much needed preliminary investigation that can contribute to the growing body of research on the syntax-prosody interface and provide new insights by reexamining the predicitons of syntactic theories with empirical data in naturalistic speech. Also, within the current dataset, we acknowledge that there were more utterances without trace than with trace. Nonetheless, we maintained confidence in our statistical analyses, as multinomial logistic regression is a versatile approach known for its robustness even in the presence of imbalanced sample sizes [14].

Apart from lenition, we found instances of two speakers inserting a filled pause between "expect" and "to" only in the utterances with an intervening trace. Filled pauses were not attested in utterances without intervening traces, even though they constituted the majority of our dataset. The filled pauses may indicate a resumption of syntactic trace between "expect" and "to." Regrettably, such instances were infrequent, so we lacked statistical power to test the association between filled pauses and intervening traces.

5. References

[1] G. Lakoff, "Global Rules," *Language*, vol. 46, no. 3, pp. 627–639, Sep. 1970, doi: 10.2307/412310.

[2] G. Goodall, "Contraction," in *The Wiley Blackwell Companion to Syntax*, 2nd ed., M. Everaert and H. C. van Riemsdijk, Eds., Hoboken, NJ: John Wiley & Sons, Inc., 2017, pp. 1–20.

[3] M. Liberman, "The distribution of wanna-contraction," Lecture notes of LING 620, University of Pennsylvania.

[4] M. Liberman, "Toward Progress in Theories of Language Sound Structure," in *Shaping Phonology*, D. Brentari and J. L. Lee, Eds., Chicago: University of Chicago Press, 2018, pp. 201–220.

[5] D. Jurafsky, A. Bell, and C. Girand, "The role of the lemma in form variation," in *Laboratory Phonology 7*, C. Gussenhoven and N. Warner, Eds., Berlin/NY: De Gruyter Mouton, 2002, pp. 3–34. doi: 10.1515/9783110197105.1.3.

[6] J. Yuan and M. Liberman, "Speaker identification on the SCOTUS corpus," *Journal of the Acoustical Society of America*, vol. 123, no. 5, p. 3878, 2008.

[7] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." 2019. [Online]. Available: http://www.praat.org/

[8] Y. Xu, "ProsodyPro -- A Tool for Large-scale Systematic Prosody Analysis," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 2013, pp. 7–10.

[9] R Core Team, "R: A Language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: https://www.r-project.org/

[10] R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*. Prentice-Hall, 2002.

[11] J. Cohen, *Statistical power analysis for the behavioral sciences*. Florence: Taylor & Francis Group, 1988.

[12] C. F. Fey, T. Hu, and A. Delios, "The Measurement and Communication of Effect Sizes in Management Research," *Management and Organization Review*, vol. 19, no. 1, pp. 176–197, Feb. 2023, doi: 10.1017/mor.2022.2.

[13] W. N. Venables and B. D. Ripley, *Modern applied statistics with S*, 4th ed. NY: Springer, 2002.

[14] A. Bayaga, "Multinomial Logistic Regression: Usage and Application in Risk Analysis," *Journal of applied quantitative methods*, vol. 5, no. 2, pp. 288–297, 2010.

[15] A. Fritz, T. Scherndl, and A. Kühberger, "A comprehensive review of reporting practices in psychological journals: Are effect sizes really enough?," *Theory Psychol*, vol. 23, no. 1, pp. 98–122, Feb. 2013, doi: 10.1177/0959354312436870.

[16] E. Elfner, "The syntax-prosody interface: current theoretical approaches and outstanding questions," *Linguistics Vanguard*, vol. 4, no. 1, p. 20160081, Jan. 2018, doi: 10.1515/lingvan-2016-0081.

[17] R. Bennett and E. Elfner, "The Syntax-Prosody Interface," *Annu Rev Linguist*, vol. 5, no. 1, pp. 151–171, Jan. 2019, doi: 10.1146/annurev-linguistics-011718-012503.

[18] J. Tanner, M. Sonderegger, and M. Wagner, "Production planning and coronal stop deletion in spontaneous speech," *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 8, no. 1, p. 15, Jun. 2017, doi: 10.5334/labphon.96.

[19] M. Wagner, "Locality in phonology and production planning," *McGill working papers in linguistics*, vol. 22, no. 1, pp. 1–18, 2012.

[20] D. G. Dunagan and M. E. L. Renwick, "Word-boundary palatalization and production planning in UK English," in *Proceedings of Meetings on Acoustics*, Acoustical Society of America, 2020, p. 060005. doi: 10.1121/2.0001394.

[21] P. M. Postal and G. K. Pullum, "The Contraction Debate," *Linguistic Inquiry*, vol. 13, no. 1, pp. 122–138, 1982.

[22] N. Chomsky, *The Minimalist Program*. Cambridge, Mass.: MIT Press, 1995.

[23] A. Radford, *Analysing English Sentences: A Minimalist Approach*. Cambridge: Cambridge University Press, 2009.

[24] A. Radford, *Syntactic Theory and the Structure of English: A Minimalist Approach*. Cambridge: Cambridge University Press, 1997.