Syntactic pausing? Re-examining the associations

Naomi Peck and Laura Becker

2022-12-14

> The right word may be effective, but no word was ever as effective as a rightly timed pause.
>
> – Mark Twain

1 Introduction

Traditional approaches to typology focus on the spoken or written word and what they can tell us about the structure of language cross-linguistically. Pauses, or the temporary 'breaks' in the flow of the language signal, receive comparatively less attention (Kirsner et al. 2002, 52). In this study, we look at the distribution of silent pauses within existing multi-language corpora to see whether their location and length correlate with any higher-level syntactic groupings in the given languages, and what this can tell us about motivations for pausing phenomena.

We know little about when, where, and why pauses occur. As we find pauses which are either filled ('um', 'uh') or unfilled (silence) in both speech and sign, we cannot simply argue that pausing solely comes from the physiological need to breathe; rather, pausing "operates in happy synchrony with some basic functional segmentations of discourse" (Chafe 1994, 57). In this study, we concentrate on silent pauses for two reasons. Firstly, theoretical claims about pauses and their relation to linguistic structure have focussed on silent pauses (Rochester 1973). Furthermore, the dataset used in this study allows for the reliable detection of silent pauses, while filled pauses were not systematically annotated in the corpora. For this reason, we use 'pauses' in the rest of the article to denote 'silent pauses', unless otherwise specified.

Traditional accounts of pausing claim that pauses occur at the juncture of two higher-level syntactic units, such as clauses (Goldman-Eisler 1968; Cooper and Paccia-Cooper 1980). Later research suggests the relevant junctures in question are in fact between prosodic units, rather than syntactic units (Gee and Grosjean 1983; Ferreira 1993), and that semantics may additionally play a role in determining the presence of a pause, with silent pauses more likely to occur between 'heavier' semantic units (O'Connell, Kowal, and Hörmann 1969; Ferreira 1993).

A number of researchers have previously attempted to categorise pauses according to their function in discourse. For example, Levelt (1993) argues that clause-/IU-internal pauses like hesitations typically reflect a word-searching task or a speech error, while clause-/IU-external pauses are often used for interactional purposes, such as holding the floor during a conversation, and give speakers a chance to process previous utterances and plan for future utterances (see also Kircher et al. 2004). Other evidence speaks to a non-unity of pause types. Studies have shown that aphasic patients show different pausing behaviour to a non-aphasic control group (Quinting 1971). Kirsner et al. (2002) take this as evidence for the existence of at least two types of pause. Moreover, Campione and Véronis (2002) compared the distribution of pause duration of read and spontaneous speech in French. They found a bimodal distribution in read speech (brief vs. medium), but a trimodal distribution in spontaneous speech (brief vs. medium vs. long). This suggests that long pauses likely result from planning needs which are absent in reading tasks, providing evidence for more than one type of pause (also cf. Goldman-Eisler 1968). Pauses are additionally argued to index social factors like speech community and social meanings such as status and hesitancy in different varieties of English (Kendall 2013; Schleef 2021). Despite the importance of the pause for successful communication, research is still largely in its infancy.

The relative lack of interest in researching pauses has left a number of outstanding empirical issues. Firstly, pauses are rarely investigated outside of major world languages. The majority of studies have been focused on European languages, with some initial studies being carried out in Mandarin Chinese (Tseng and Chang 2008; Fon, Johnson, and Chen 2011). Furthermore, most of these studies have been conducted in lab conditions using either reading tasks or controlled stimuli. Most of these studies operationalise pausing as silences of 250ms or longer in the speech signal and do not include filled pauses in their models (Rochester 1973). While this method of data collection is valuable for collecting psycholinguistic data on the role of the pause, the lack of work with spontaneous data means that it is difficult to generalise these insights to the level of discourse. Lastly, very little work has been done into the association between pause location and pause duration. More empirical work on pausing can also help to shed more light on psycholinguistic aspects of language production.

In our study, we investigated how pauses were distributed across 8 typologically distinct languages in the Multi-CAST corpora (Haig and Schnell 2021). The languages are listed in Table 1 together with information on the number of different speakers in the corpus, the number of utterances, and the genres of texts used. We chose a subset of data which was consisted of realistic monologues, i.e. autobiographical narratives (AN) and traditional narratives (TN) at the time of writing. Only

Tondano includes one additional stimulus-based narrative (SN). As can be seen in Table 1, each corpus ranges from 160 to 1300 utterances per language from 2 to 6 different speakers. We used the provided .wav files, along with the accompanying GRAID annotations in .eaf format (Haig and Schnell 2014).

Table 1 : Overview of dataset

language	family	speaker	utt	genre	reference
Arta	Austronesian	3	227	AN, TN	(Kimoto 2019)
Nafsan	Austronesian	3	163	TN	(Thieberger and Brickell 2019)
Teop	Austronesian	4	1019	TN	(Mosel 2015)
Tondano	Austronesian	6	1254	SN, AN	(Brickell 2016)
Mandarin	Sino-Tibetan	3	844	TN	(Vollmer 2020)
Sanzhi Dargwa	Nakh-Daghestanian	3	294	AN, TN	(Forker and Schiborr 2019)
Tabasaran	Nakh-Daghestanian	2	629	TN, AN	(Bogomolova 2021)
N. Kurdish	Indo-European	2	555	TN	(Haig, Vollmer, and Thiele 2015)

Using this dataset, we investigate three research questions:

1. If there is a pause, is there a clause boundary?
2. How strongly are different clausal contexts associated with pauses?
3. Does the clausal context affect the duration of a pause?

2 Methodology & data processing

The majority of language documentation corpora are constructed for grammatical research questions, and as such, rarely have pre-existing notations for non-morphosyntactic phenomena. The re-use of these databases for phonological and/or prosodic research questions thus necessitates another round of annotation by researchers. In our case, this required the annotation of silent pauses, a task which has been semi-automated in the past.

Periods of silence in an audio signal can be extracted automatically using a silence recogniser. This does not necessarily capture all and only meaningful silent pauses, but reduces the manual labour required by an annotator in the process of marking pauses (see also San et al. 2022). We followed the instructions by Kashima (2017) in using Praat to automatically annotate periods of silence and speech using the "to TextGrid (Silences)" function. We used the suggested pitch-related settings (pitch = 70Hz, silence threshold = -35dB); however, we reduced the minimum silent and sounding interval durations to 50ms and 100ms respectively to capture potentially meaningful pauses under 100ms (cf. Campione and Véronis 2002; Kendall 2013). These thresholds ensure that short bursts of noise are automatically excluded, such as common noises in the fieldwork context like claps used to synchronise audiovisual files. The results were then saved as TextGrids and imported into the respective ELAN file as a separate tier. As the relatively low silence threshold also results in the capture of some phonetic pauses such as plosive holds, the segmentation was later manually corrected in ELAN ("ELAN" 2022). We additionally deleted intervening utterances made by other speakers. The exclusion of such utterances ensured that we only included utterances from one speaker per file, and that the data was kept as monologic as possible.

After the manual annotation in ELAN, we exported the data into a spreadsheet format for further automatic processing in R (R Core Team 2020). Figure 1 shows the first elements from the texts "nafsan_tafra" and "nafsan_nmatu" in the exported but not yet fully processed spreadsheet format. Each row represents speech or silent pause segments by grammatical word. For better readability, the rows containing silent pauses are shaded gray in Figure 1. In the case a grammatical word falls on the boundary between speech and pause segments due to the original alignment in ELAN, it is repeated in separate rows as speech and pause, e.g., *teetwei* in rows 14 and 15. Each row corresponds to a grammatical word split into silent pause or speech segments, and is assigned a begin and end time (in ms) as well as the resulting duration. In addition, the annotation contains information on the language, the text, the utterance id, the grammatical word, its GRAID annotation, and gloss. We also added the speaker information to the extracted data using the *multicastR* package (Schiborr 2021).

Based on the GRAID and gloss information, we added annotations for clause boundaries. We distinguished between main clause boundary (`cb_main`), dependent clause boundary (`cb_dep`) and clause-internal boundary (`no_cb`). A GRAID annotation of "##" was marked as a main clause boundary, and a gloss annotation of "#" or "%" was coded as a dependent clause boundary. All other remaining rows were analysed as having no clause boundary, i.e. as being clause-internal

N	begin	end	duration	graid	silences	gloss	gram_words	utterance_id	language	text	cb	speaker
1	400	630	230	##	pause	#	#	nafsan_tafra_0001	nafsan	tafra	cb_main	NF03
2	630	878	248	##	speech	#	#	nafsan_tafra_0001	nafsan	tafra	cb_main	NF03
3	878	1356	478	#ac	speech	#	#	nafsan_tafra_0001	nafsan	tafra	cb_dep	NF03
4	1356	1835	479	other	speech	while	selwan	nafsan_tafra_0001	nafsan	tafra	no_cb	NF03
5	1835	2313	478	0.1:a	speech	ZERO	ZERO	nafsan_tafra_0001	nafsan	tafra	no_cb	NF03
6	2313	2791	478	"=lv"	speech	1P.inRS=	tu=	nafsan_tafra_0001	nafsan	tafra	no_cb	NF03
7	2791	3270	479	v:pred	speech	arrive	paakor	nafsan_tafra_0001	nafsan	tafra	no_cb	NF03
8	3270	3560	290	np:p	speech	end	nametp̃ag	nafsan_tafra_0001	nafsan	tafra	no_cb	NF03
9	3560	3748	188	np:p	pause	end	nametp̃ag	nafsan_tafra_0001	nafsan	tafra	no_cb	NF03
10	3748	4226	478	rn_np	pause	year	ntau	nafsan_tafra_0001	nafsan	tafra	no_cb	NF03
11	4226	4705	479	%	pause	%	%	nafsan_tafra_0001	nafsan	tafra	cb_dep	NF03
12	4705	5183	478	0.1:a	pause	ZERO	ZERO	nafsan_tafra_0001	nafsan	tafra	no_cb	NF03
13	5183	5350	167	"=lv"	pause	1D.exRS=	ra=	nafsan_tafra_0001	nafsan	tafra	no_cb	NF03
	...	...	...	...	...	...	...	...	...	...	...	...
14	16876	17030	154	other	speech	before	teetwei	nafsan_nmatu_0001	nafsan	nmatu	no_cb	NF03
15	17030	17400	370	other	pause	before	teetwei	nafsan_nmatu_0001	nafsan	nmatu	no_cb	NF03
16	17740	18184	444	##	pause	#	#	nafsan_nmatu_0002	nafsan	nmatu	cb_main	NF03
17	18184	18366	182	##	speech	#	#	nafsan_nmatu_0002	nafsan	nmatu	cb_main	NF03
18	18366	18993	627	0:a	speech	ZERO	ZERO	nafsan_nmatu_0002	nafsan	nmatu	no_cb	NF03
19	18993	19620	627	"=lv"	speech	3S.RS=	i=	nafsan_nmatu_0002	nafsan	nmatu	no_cb	NF03

Figure 1: Example of an exported sentence from Nafsan in the spreadsheet format.

boundaries.

The next processing steps were to filter out rows that did not belong to any utterance (additional comments or background utterances not belonging to the monologue) and to exclude the first and last silent segment of each text, i.e. silent elements at the beginning and at the end of the monologue. For the example sentence shown in Figure 1, this entails the manual deletion of row 1. In addition, we deleted all rows that are annotated as text-initial clause boundaries, as those do not have any preceding context that could give rise to a silent pause. Therefore, rows 2 and 3 from the example in Figure 1 were deleted as well. In those cases in which the rows of text-initial clause boundaries were marked as speech, as in rows 2 and 3, we added their speech duration to the next row annotated as `no_cb`, here row 4. The value of the "begin" column for row 4 in 1 was thus changed to 630ms from 1356ms.

As can be seen in Figure 1, the pre-processed output contains speech and pause segments that are split up by grammatical words. We therefore grouped words according to their co-occurrence with pauses or speech and created single annotations for each group. Each segment was then automatically annotated with the total duration of the grouped words. This led to the identification of, for instance, a single pause segment from rows 9 to 13 with the duration of 5350ms - 3560ms = 1790ms, and a single speech segment from rows 2 to 8 with the duration of 3560ms - 630ms = 2930ms.

Our next step was to identify pause segments which co-occurred with clause boundaries. When a clause boundary fell within a pause group, like the dependent clause boundary in row 11, we automatically annotated the pause with the according clause boundary type.[1] Due to the alignment of the acoustic signal and the original annotations, a word annotation containing a clause boundary sometimes fell across the edge between a pause and a speech segment. This resulted in the output of two separate rows containing a clause boundary, as in rows 16 and 17 in Figure 1. In such cases, we deleted the row containing a clause boundary falling within a speech segment, i.e. row 17, to avoid the artificial creation of clause boundaries that occur without silent pauses (given that the same clause boundary does in fact fall on a pause segment). The duration of the second annotation in row 17 was added to the following speech segment. After processing, the resulting main clause boundary in row 16 received a combined pause length of 18184ms - 17030ms = 1154ms.

The resulting processed data annotated for speech and pause segments, their lengths, and their occurrence at clause boundaries or within clauses can be found in the supplementary materials. Figure 2 shows the extract from Figure 1 in its final format as used for the analysis.

N	silences	utterance_id	language	text	cb	speaker	duration
1	speech	nafsan_tafra_0001	nafsan	tafra	no_cb	NF03	2930
2	pause	nafsan_tafra_0001	nafsan	tafra	cb_dep	NF03	1790
	...	...	...	...	...	...	...
3	speech	nafsan_nmatu_0001	nafsan	nmatu	no_cb	NF03	2772
4	pause	nafsan_nmatu_0001	nafsan	nmatu	cb_main	NF03	1154
5	speech	nafsan_nmatu_0002	nafsan	nmatu	no_cb	NF03	1547

Figure 2: Example of a processed sentence from Nafsan.

3 Results

3.1 Overall Tendencies

Before examining the association between silent pauses and different types of clausal contexts, Table 2 shows how the three clausal contexts are distributed in the 8 languages. As dependent clauses can be more or less frequent in different languages, this entails varying proportions of clause boundaries for each language. This in turn influences the distribution of clausal contexts across silent pauses discussed in Section 3.2.

Table 2 shows that there are clear differences in the amount of dependent clauses used despite the fact that the proportion of annotations with no clause boundary is roughly similar across languages.

[1]When a silent pause segment contained both a main and a dependent clause boundary, we annotated it as a main clause boundary, given that this type is the "stronger" boundary type. This is in line with how GRAID operationalises the co-occurrence of the end of a dependent clause and the beginning of a new main clause, in that the end of the dependent clause should not be annotated (Haig and Schnell 2014, 23).

Table 2: Proportion of clausal contexts.

Language	no_cb	cb_main	cb_dep
Arta	0.67	0.14	0.19
Mandarin	0.62	0.24	0.14
Nafsan	0.61	0.23	0.16
Northern Kurdish	0.56	0.09	0.35
Sanzhi Dargwa	0.60	0.12	0.28
Tabarasan	0.58	0.12	0.30
Teop	0.58	0.24	0.18
Tondano	0.70	0.24	0.06
Average	0.61	0.18	0.21

For example, `cb_dep` annotations make up 35% of the total annotations in Northern Kurdish, while `cb_main` annotations make up only 9%.

Figure 3 builds on the distribution of clausal contexts by showing the distribution of no clause boundaries (orange), main (blue) and dependent (green) clause boundaries across speech and silent pause segments in the 8 languages. As expected, the majority of speech segments occurs within clauses (i.e. `no_cb`) in all languages. However, we do also find both main and dependent clause boundaries that occur within speech segments, meaning that they are not accompanied by a silent pause.

Comparing the distribution of the three clausal contexts in pause and speech segments (left vs. right bars), we see clearly that pause segments are made up of a higher proportion of clause boundaries than speech segments in all languages. However, the exact proportions differ across languages, with clausal boundaries making up 75% of all pauses in Northern Kurdish but only around 30% of all pauses in Tondano. Both patterns are at least partially a consequence of the proportion of clausal boundaries in the data itself. As Table 1 showed, clausal boundaries make up 44% of all boundary context in the Northern Kurdish data. It is therefore not surprising that we also find a high proportion of clausal boundaries in the pause segments. Tondano shows the opposite pattern, with the lowest overall proportion of clausal boundaries (30%). This distribution could account for the low proportion of clausal boundaries in the pause segments shown in Figure 3.

These raw distributions of clausal contexts across speech and pause segments indicate that the association between clause boundaries and pauses is more complex than a simple one-to-one relationship. Clause boundaries need not be accompanied by silent pauses and pauses also frequently occur within clause boundaries. This suggests that a closer look at the co-occcurrence and distribution patterns of silent pauses and clausal contexts is indeed warranted.

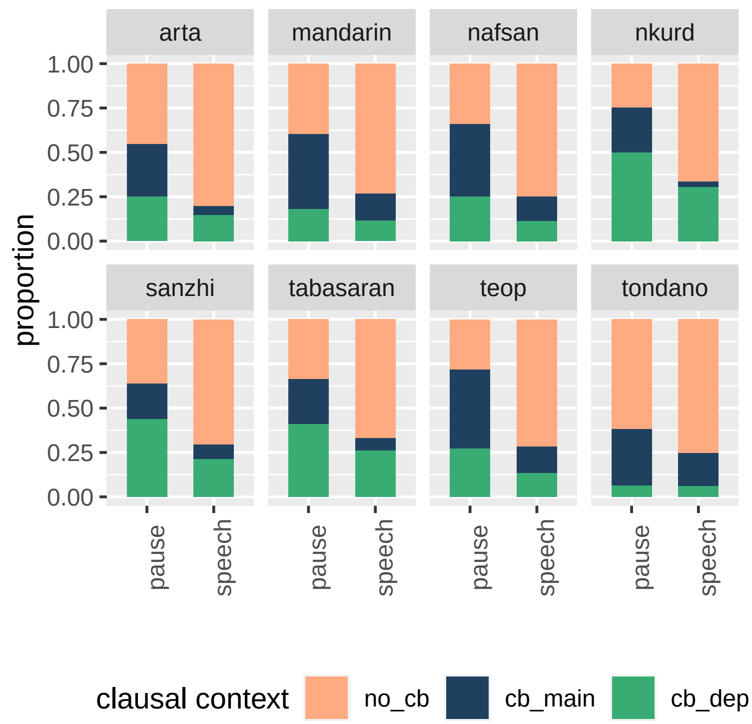

Figure 3: The raw distribution of clausal contexts across speech and pause segments.

3.2 The distribution of clausal contexts across silent pauses

To investigate the co-occurrence and distribution of silent pauses and clausal contexts, we extracted all datapoints that correspond to silent pauses. We then fitted a multinomial regression model to assess the probability of pauses occurring with the three clausal contexts (`no_cb`, `cb_dep`, `cb_main`) using the languages as a population level effect and single speakers as a group-level effect (cl_context ~ 1 + language + (1|speaker)). The model was fitted using Bayesian methods with Stan (Carpenter et al. 2017) and brms (Bürkner 2017) in R (R Core Team 2020).[2]

Figure 4 shows the conditional effects of the fitted model. This model predicts the probabilities of clausal contexts in silent pause segments across languages, taking into account variation across single speakers. The points in Figure 4 correspond to the mean of the probability density of the predictions; the whiskers correspond to the 95% uncertainty intervals. This means that we can be 95% confident that the actual probability will lie within this interval based on the data and the model.

The model results in Figure 4 confirm the raw distributions shown in Section 3.1. We find considerable variation across languages as to what proportion of silent pauses falls onto clause boundaries.

[2]The code for all models can be found in the supplementary materials.

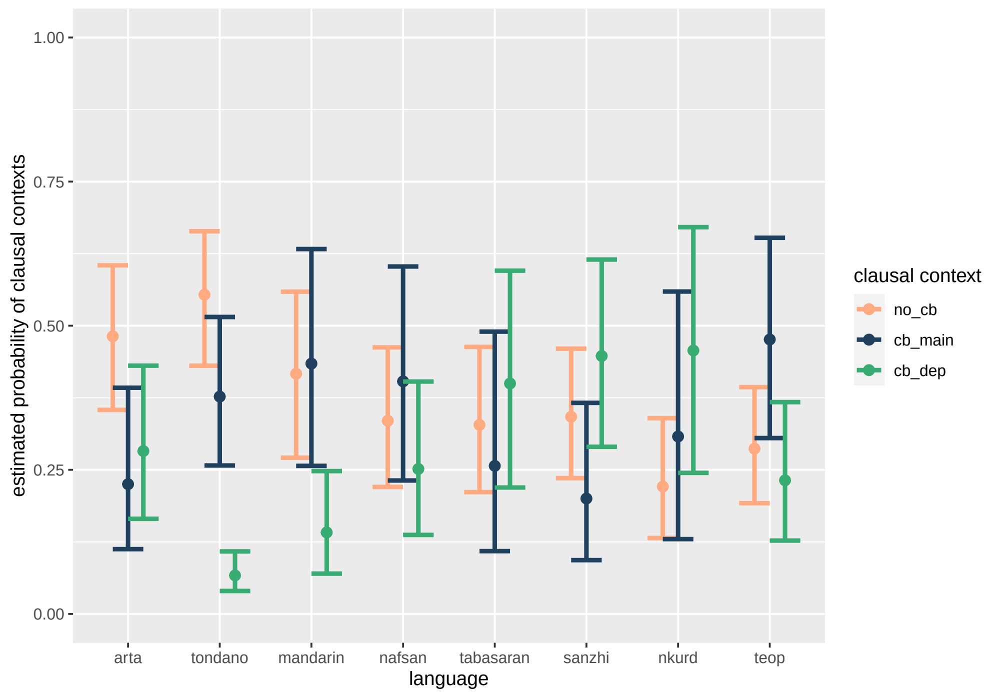

Figure 4: Predicted probablity of clausal contexts in silent pauses.

The only language where pauses appear to be more likely within clauses than at any type of clause boundary is Arta. In Tondano and Mandarin, pauses are most likely to occur within clauses or at main clause boundaries, and the probability of pauses at dependent clause boundaries is substantially lower. This is at least partially due to the overall low proportion of dependent clause boundaries in those two languages to begin with (cf. Figure 2). In Nafsan, Tabasaran, Sanzhi Dargwa, and to a certain extent in Northern Kurdish, pauses are fairly equally distributed across the three syntactic contexts. Teop is the only language where there is a slight preference for pauses to occur at main clause boundaries over the other two contexts. Interestingly, the preference for silent pauses at main clause boundaries in Teop is not due to a low proportion of dependent clause boundaries in the language. Rather, it appears to represent a genuine preference for pausing in this context.

3.3 The associations of clause boundaries with pauses

Another perspective of looking at the associations between clause boundaries and pauses is to examine the probability of the co-occurrence of pauses with the three different clausal contexts. To do this, we used the full dataset including speech and silent pause segments to fit a Bayesian logistic regression model to assess the probability of pauses across the clausal contexts. We added varying intercepts as well as varying slopes over clausal contexts for speakers (pause ~ 1 + clause_context

* language + (1 + clause_context|speaker)). Figure 5 shows the predicted probabilities of pauses across the contexts of no clause boundary, dependent, and main clause boundaries.

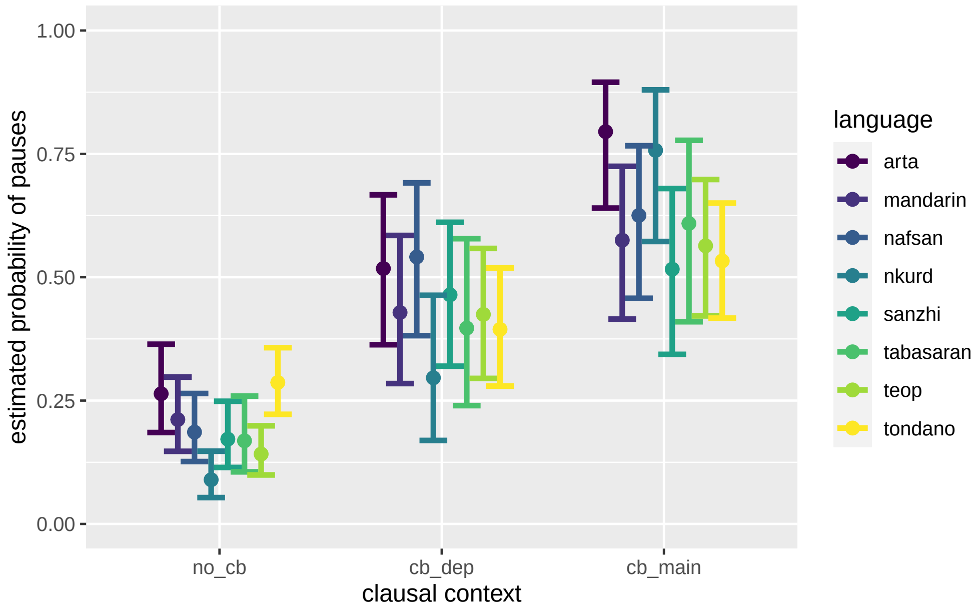

Figure 5: Predicted probablity of pauses across clausal contexts.

While we see a certain degree of variation across the 8 languages, the overall pattern is robust. Both main and dependent clause boundaries have a slightly higher probability of silent pauses than no clause boundary positions. For main clause boundaries, the probability of co-occurrence with pauses is above 0.5 in all languages, with a probability above 0.75 in Arta and Nafsan. For dependent clause boundaries, the predicted co-occurrence probabilities for all languages are slightly lower, ranging from 0.3 to just above 0.5. For the no clause boundary context, we find the overall lowest probability of pause co-occurrence, from 0.1 to 0.3.

We generally find a higher level of cross-linguistic variation in the `cb_main` and `cb_dep` contexts than in the `no_cb` condition. However, the variation across languages is consistent to a certain extent. For example, Arta generally has a higher predicted probability of pauses than Teop across all three contexts. If the occurrence of pauses was solely associated with high-level syntactic units, we would expect the model to predict a close-to-zero probability of pauses in the no clause boundary condition. Even though our results point to a rather low probability of pauses occurring without a clause boundary, pauses do occur without clause boundaries in our dataset. This supports the view that there is no direct relation between pausing and syntax.

The robust patterning of pause co-occurrence probabilities across our dataset illustrates a clear trend. As the level of the syntactic boundary juncture increases, so too does the overall predicted probability of a pause co-occurring. This suggests that there is no categorical distinction either between main clause boundaries and other contexts (`cb_main` vs. `cb_dep` and `no_cb`) or between non-clausal boundaries and clausal boundaries (`no_cb` vs. `cb_dep` and `cb_main`). Rather, the probability of pause co-occurrence appears to be gradient in nature.

3.4 Pause durations

Previous work has suggested that pauses not only differ in their distribution with respect to clausal boundaries, but that they also show differences in length (e.g. Krivokapić 2007). Figure 6 shows the raw distribution of pause durations across the 8 languages of the dataset.

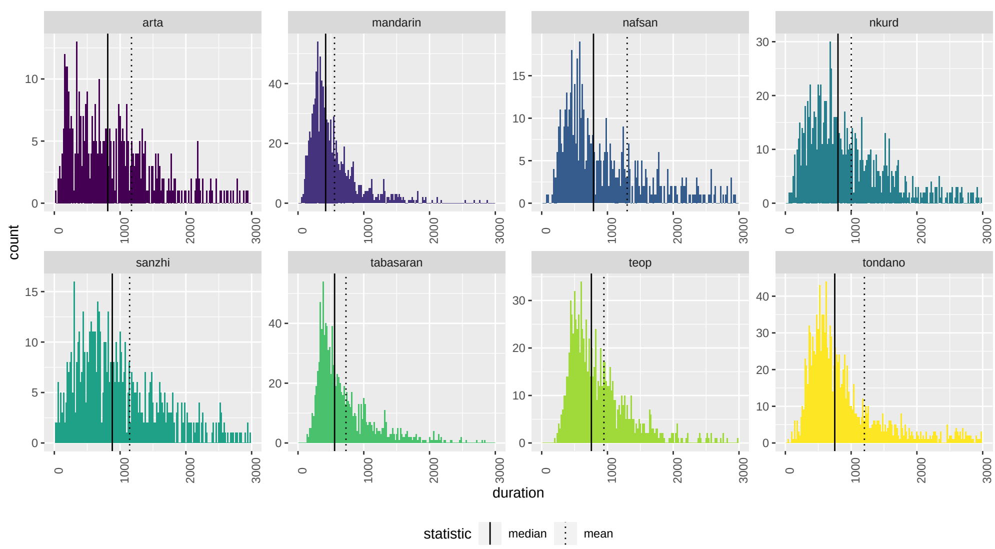

Figure 6: Observed pause duration by language.

The general trends are fairly robust. The majority of silent pauses in the dataset have a length of < 750ms. Only Arta, Northern Kurdish and Sanzhi Dargwa have a somewhat larger portion of silent pauses longer than 750ms. Few pauses are longer than 10000ms and pauses longer than 25000ms are rare.

The average pause duration is more or less comparable across languages. As can be seen in Figure 6, the median pause duration (the solid vertical line) is between 500ms and 1000ms for all languages, save for Mandarin, which has a slightly shorter average pause duration. However, we see more variation of the mean pause duration (dotted black line) across languages. Arta, Nafsan, and Tondano evidence a larger difference between median and mean pause duration, as these three languages have more longer pauses than the other languages of the dataset. We can thus say that the distribution of pause length, albeit not identical, is comparable across the 8 languages of the dataset.

Figure 7 shows the distribution of pause duration together with median and mean values across the three clausal contexts: no clause boundary, main and dependent clause boundaries. These raw distributions together with the median and mean suggest a difference in pause duration associated with the absence or presence of a clause boundary. Most pauses in the `no_cb` context have a length of below 750ms, with only few pauses that are longer than 750ms and even fewer above 1000ms. However, we find quite a substantial proportion of pauses with durations above 1000ms up to 3000ms for both clause boundary contexts. The difference in pause duration between no clause boundary and clause boundary contexts is further reflected in their median pause durations: 543ms for no clausal boundaries vs. 800ms and 825ms for dependent and main clause boundaries, respectively.

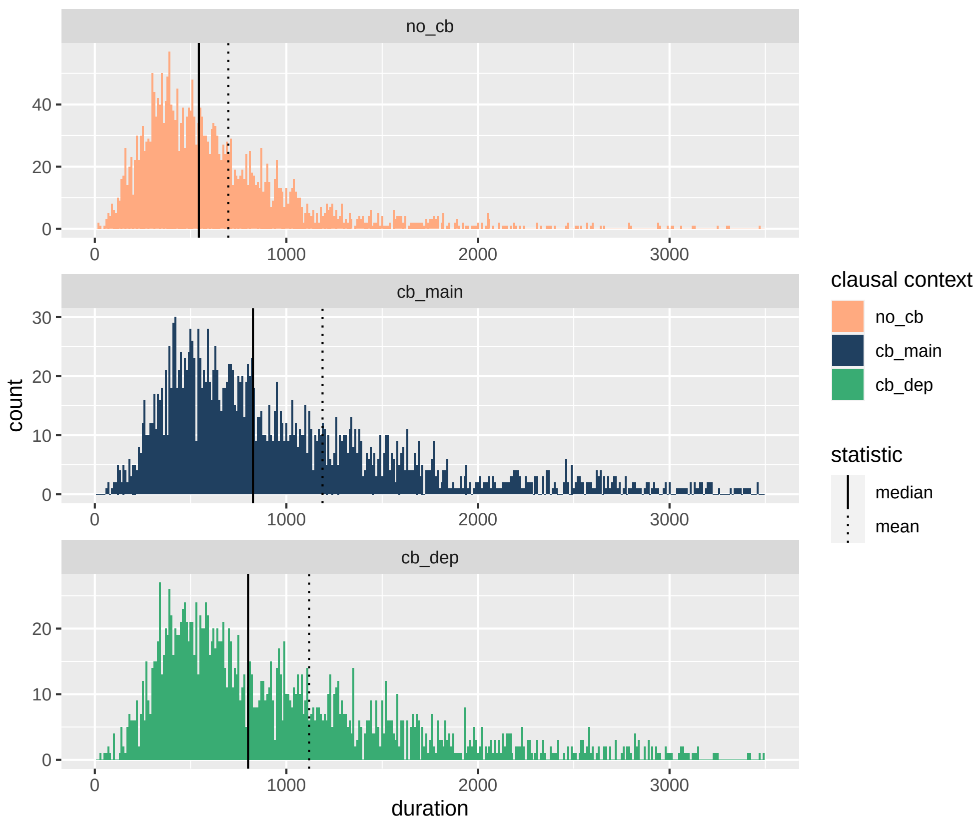

Figure 7: Observed pause duration by clausal context.

As such, the distributions of pause duration point to a two-way distinction of pauses between clause boundary positions and positions within clauses. To test this generalisation, we fitted a Bayesian regression model to predict the duration of pauses from clausal contexts and languages while controlling for the effects of single speakers: pause_duration ~ clausal_context * language + (1 + clausal_context|speaker). The conditional effects of clausal contexts and languages on pause durations are shown in Figure 8.

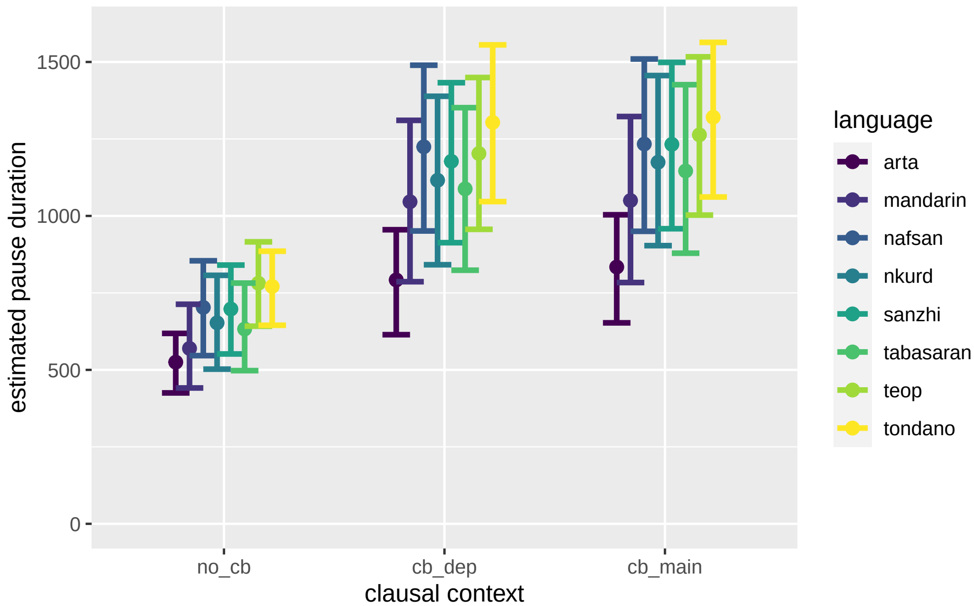

Figure 8: Estimated pause duration by clausal context.

The model predictions shown in Figure 8 support the hypothesis of a two-way distinction of pauses based on their duration. Pauses are predicted to be between 520ms and 770ms very consistently across languages when no clause boundary is present. At main and dependent clause boundaries, the average predicted pause duration varies to a greater extent across languages. Pauses are predicted to be between 1000ms and 1500ms for both types of clausal boundaries.

While most languages in our dataset pattern similarly with regards to estimated pause duration, Arta exhibits shorter pause durations in all three contexts compared to the other languages in Figure 8. However, the relative difference between the no clausal boundary and clausal boundary contexts is maintained. Given that pause duration is heavily speaker-dependent (Goldman-Eisler 1968), we cannot conclude that this is a real crosslinguistic difference. Rather, this may be the result of idiosyncratic pause durations of single speakers in the dataset.

The overall results concerning pause duration strongly point to a two-way difference of pause duration associated with the absence or presence of clause boundaries. Pauses at clause boundaries tend to be longer than pauses not at clause boundaries across languages. Unlike the probability of the presence of a pause (cf. Section 3.3), the type of clause boundary, i.e. main vs. dependent, does not seem to play an important role for pause duration.

4 Discussion

The results of our study evidence a complex interaction between pausing and clause boundaries. Pauses occur both within clauses and at clause boundaries without any strong cross-linguistic preference, as suggested by the very large uncertainty intervals in Figure 4. This suggests that the traditional assumption that the occurrence of silent pauses is associated with certain syntactic positions such as clause boundaries has to be called into question.

Moreover, the distribution of pauses across clause contexts (Figure 5) appears to be language-specific. For example, Arta has a comparatively high estimated probability of pauses at main clause boundaries and at dependent clause boundaries. Teop tends to have generally a lower predicted probability of pauses. For the other languages, the patterns are less straightforward. For instance, Northern Kurdish shows a relatively low probability for pauses in the `no_cb` and `cb_dep` contexts, but a higher probability for pauses in the `cb_main` context compared to the other languages. This is remarkable due to the low proportion of main clause boundaries in the dataset (Table 2), likely because of a higher proportion of dependent clause contexts in the dataset. That is to say, main clause boundaries are less frequent to begin with, but are more likely to be accompanied by a pause, suggesting that there is a stronger association between main clause boundaries and pauses in Northern Kurdish. From this follows that the typological profile of a language can impact the distribution of pauses.

Pause duration appears to be more stable cross-linguistically. This is remarkable because it has been shown that pause duration is heavily speaker-dependent (Goldman-Eisler 1968). Despite overall tendencies being similar across languages in the dataset, Figure 8 showed that pause durations in Arta are shorter. This could be due to idiosyncratic behaviour of the speakers rather than a real difference between languages. Teasing apart these issues will be important to understand the factors which affect pause duration.

Previous studies have suggested that there is more than one type of pause. Campione and Véronis (2002) argued for three types of pauses based on duration: brief (< 200 msecs), medium (200-1000 msecs), and long (> 1000 msecs). While the aim of this study was to investigate associations between pauses and syntactic contexts, our results support the distinction between medium and long pauses based on the categorical difference of clause boundary and no clause boundary contexts. The estimated pause duration in the no clause boundary context (Figure 8) corresponds to the medium category, while the duration in the clause boundary contexts maps onto the long pause category. This two-way distinction also echoes psycholinguistic work (Levelt 1993; Kirsner et al. 2002). More work is needed to ascertain to what extent differences in duration reflect different pause types.

5 Conclusion

This study supports previous studies which argue against a direct relation between syntax and prosody. Our naturalistic and linguistically-diverse sample showed that pauses were not strongly associated with clause boundaries and were, in fact, more likely to occur within clauses. Clause boundaries were associated with pauses, with stronger boundaries more likely to co-occur with a pause. The results from pause duration points to a categorical distinction between co-occurrence with clause boundaries and within clauses. Apart from investigating pausing, our study showed how corpora built for other purposes can be used in the investigation of other research questions.

A number of improvements could be made to gain further insight into the role of pauses. As pausing is not central to discussions of linguistic structure, transcriptions of naturalistic speech have focussed less on faithfully representing both filled and silent pauses. More detailed transcriptions of pause-related phenomena will help to better understand the role of pausing in language. The control of extralinguistic factors such as speakers and text genre could additionally be improved, especially given the sociolinguistic importance of the pause. Further refinement of our approach to pausing could also consider other factors which affect pause location, rate, and duration, such as the inclusion of information on other prosodic, syntactic, and semantic units, as well as speech rate. An important further step would be to include conversational data to examine the interaction of pausing and turn-taking.

Acknowledgements

We would like to thank both Wifek Bouaziz and Sharon Kromer for their work in helping us manually align the data. A further thanks to those who gave enthusiastic feedback about this project at our departmental reading group at Freiburg and at the Language Documentation and Theory 6 workshop on intonation.

References

Bogomolova, Dmitry & Schiborr, Natalia & Ganenkov. 2021. "Multi-CAST Tabasaran." In *Multi-CAST: Multilingual corpus of annotated spoken texts*, edited by Geoffrey Haig and Stefan Schnell.

Brickell, Timothy. 2016. "Multi-CAST Tondano." In *Multi-CAST: Multilingual corpus of annotated spoken texts*, edited by Geoffrey Haig and Stefan Schnell.

Bürkner, Paul-Christian. 2017. "brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80 (1): 1–28. https://doi.org/10.18637/jss.v080.i01.

Campione, Estelle, and Jean Véronis. 2002. “A Large-Scale Multilingual Study of Silent Pause Duration.” In *Proceedings of Eurospeech 2002*, 199–202.

Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. “Stan: A Probabilistic Programming Language.” *Journal of Statistical Software* 76 (1): 1–32. https://doi.org/10.18637/jss.v076.i01.

Chafe, Wallace L. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.

Cooper, William E., and Jeanne Paccia-Cooper. 1980. *Syntax and Speech*. Cambridge, MA: Harvard University Press.

“ELAN.” 2022. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive.

Ferreira, Fernanda. 1993. “Creation of Prosody During Sentence Production.” *Psychological Review* 100 (2): 233–53.

Fon, Janice, Keith Johnson, and Sally Chen. 2011. “Durational Patterning at Syntactic and Discourse Boundaries in Mandarin Spontaneous Speech.” *Language and Speech* 54 (1): 5–32. https://doi.org/fj48s4.

Forker, Diana, and Nils N. Schiborr. 2019. “Multi-CAST Sanzhi Dargwa.” In *Multi-CAST: Multilingual corpus of annotated spoken texts*, edited by Geoffrey Haig and Stefan Schnell.

Gee, James Paul, and François Grosjean. 1983. “Performance Structures: A Psycholinguistic and Linguistic Appraisal.” *Cognitive Psychology* 15 (4): 411–58. https://doi.org/cbfbsj.

Goldman-Eisler, Frieda. 1968. *Psycholinguistics: Experiments in Spontaneous Speech*. London: The Academic Press.

Haig, Geoffrey, and Stefan Schnell. 2014. “Annotations Using GRAID (Grammatical Relations and Animacy in Discourse): Introduction and Guidelines for Annotators, Version 7.0.”

———, eds. 2021. “Multi-CAST: Multilingual Corpus of Annotated Spoken Texts. Version 2108.”

Haig, Geoffrey, Maria Vollmer, and Hanna Thiele. 2015. “Multi-CAST Northern Kurdish.” In *Multi-CAST: Multilingual corpus of annotated spoken texts*, edited by Geoffrey Haig and Stefan Schnell.

Kashima, Eri. 2017. “Elan/Praat Machine Segmenting.” *Yammering On*.

Kendall, Tyler. 2013. *Speech Rate, Pause and Sociolinguistic Variation*. London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137291448.

Kimoto, Yukinori. 2019. “Multi-CAST Arta.” In *Multi-CAST: Multilingual corpus of annotated spoken texts*, edited by Geoffrey Haig and Stefan Schnell.

Kircher, Tilo T. J, Michael J Brammer, W Levelt, Mathias Bartels, and Philip K McGuire. 2004. “Pausing for Thought: Engagement of Left Temporal Cortex During Pauses in Speech.” *Neu-*

roImage 21 (1): 84–90. https://doi.org/bfppft.

Kirsner, Kim, John Dunn, Kathryn Hird, Tim Parkin, and Craig Clark. 2002. "Time for a Pause..." In *Proceedings of the 9th Australian International Conference on Speech Science & Technology Melbourne, December 2 to 5, 2002*, 52–57.

Krivokapić, Jelena. 2007. "Prosodic Planning: Effects of Phrasal Length and Complexity on Pause Duration." *Journal of Phonetics* 35 (2): 162–79. https://doi.org/drvfgg.

Levelt, Willem J. M. 1993. *Speaking: From Intention to Articulation*. The MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001.

Mosel, Stefan, Ulrike & Schnell. 2015. "Multi-CAST Teop." In *Multi-CAST: Multilingual corpus of annotated spoken texts*, edited by Geoffrey Haig and Stefan Schnell.

O'Connell, Daniel C., Sabine Kowal, and Hans Hörmann. 1969. "Semantic Determinants of Pauses." *Psychologische Forschung* 33 (1): 50–67. https://doi.org/cnc7cf.

Quinting, Gerd. 1971. *Hesitation Phenomena in Adult Aphasic and Normal Speech*. The Hague: De Gruyter Mouton.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rochester, S. R. 1973. "The Significance of Pauses in Spontaneous Speech." *Journal of Psycholinguistic Research* 2 (1): 51–81. https://doi.org/fn76bg.

San, Nay, Martijn Bartelds, Tolúlọpẹ́ Ògúnrẹ̀mí, Alison Mount, Ruben Thompson, Michael Higgins, Roy Barker, Jane Simpson, and Dan Jurafsky. 2022. "Automated Speech Tools for Helping Communities Process Restricted-Access Corpora for Language Revival Efforts." *arXiv:2204.07272 [Cs, Eess]*, April. https://arxiv.org/abs/2204.07272.

Schiborr, Nils Norman. 2021. *multicastR: A Companion to the Multi-CAST Collection*. https://multicast.aspra.uni-bamberg.de/.

Schleef, Erik. 2021. "Mechanisms of Meaning Making in the Co-Occurrence of Pragmatic Markers with Silent Pauses." *Language in Society*, August, 1–27. https://doi.org/10.1017/S0047404521000610.

Thieberger, Nick, and Timothy Brickell. 2019. "Multi-CAST Nafsan." In *Multi-CAST: Multilingual corpus of annotated spoken texts*, edited by Geoffrey Haig and Stefan Schnell.

Tseng, Chiu-Yu, and Chun-Hsiang Chang. 2008. "Pause or No Pause? Prosodic Phrase Boundaries Revisited." *Tsinghua Science and Technology*, 10.

Vollmer, Maria. 2020. "Multi-CAST Mandarin." In *Multi-CAST: Multilingual corpus of annotated spoken texts*, edited by Geoffrey Haig and Stefan Schnell.