Wait for it!
Stronger influence of context on categorical perception in Danish than Norwegian

Byurakn Ishkhanyan (byurakn@cc.au.dk)
School of Communication and Culture, Aarhus University, Denmark

Anders Højen (hojen@cc.au.dk)
School of Communication and Culture, Aarhus University, Denmark

Riccardo Fusaroli (fusaroli@cas.au.dk)
School of Communication and Culture & Interacting Minds Centre, Aarhus University, Denmark

Christer Johansson (christer.johansson@uib.no)
Department of Linguistic, Literary and Aesthetic studies, University of Bergen, Norway

Kristian Tylén (kristian@cc.au.dk)
School of Communication and Culture & Interacting Minds Centre, Aarhus University, Denmark

Morten H. Christiansen (christiansen@cornell.edu)
Cornell University, Department of Psychology, Ithaca, NY 14850 USA
School of Communication and Culture & Interacting Minds Centre, Aarhus University, Denmark

Abstract

Speech input is often noisy and ambiguous. Yet listeners usually do not have difficulties understanding it. A key hypothesis is that in speech processing acoustic-phonetic bottom-up processing is complemented by top-down contextual information. This context effect is larger when the ambiguous word is only separated from a disambiguating word by a few syllables compared to many syllables, suggesting that there is a limited time window for processing acoustic-phonetic information with the help of context. Here, we argue that the relative weight of bottom-up and top-down processes may be different for languages that have different phonological properties. We report an experiment comparing two closely related languages, Danish and Norwegian. We show that Danish speakers do indeed rely on context more than Norwegian speakers do. These results highlight the importance of investigating cross-linguistic differences in speech processing, suggesting that speakers of different languages may develop different language processing strategies.

Keywords: categorical perception; speech perception; Danish; Norwegian, cross-linguistic studies

Introduction

Speech is often ambiguous and noisy. Yet most of the time listeners show remarkable skills in understanding what is being said. A possible explanation is that the imperfect acoustic-phonetic input is integrated with contextual information. Thus, to understand speech, listeners combine bottom-up acoustic-phonetic cues with top-down lexical-semantic and pragmatic contextual information. This context effect might be particularly apparent when the acoustic-phonetic information is unclear or noisy (e.g., Borsky, Tuller & Shapiro, 1998; Gaskell & Marslen-Wilson, 2001; Marslen-Wilson & Welsh, 1978; Samuel, 1981).

Despite the variability of the acoustic properties of individual sounds and the noisiness of the acoustic-phonetic input, the perception of speech sounds is *categorical* (Liberman et al., 1957). This means that a certain sound is usually perceived unambiguously (e.g., either as a /b/ or as a /p/); listeners ignore within-category acoustic differences while easily perceiving across-category acoustic differences of the same magnitude.

Both within-word and sentential context facilitate sound categorization when the acoustic-phonetic information is ambiguous (Brown-Schmidt & Toscano, 2017; Bushong & Jaeger, 2017; Connine, Blasko & Hall, 1991; McMurray, Tanenhaus & Aslin, 2009; Szostak & Pitt, 2013). In a phoneme identification study, Connine et al. (1991) manipulated the onset of the target words *dent/tent* on a continuum from a clear [d] to a clear [t^h] with three intermediate steps. The listeners were presented with sentences biased either towards *dent* (*After the _ent corroded, they patched it*) or towards *tent* (*After the _ent collapsed, we went home*). Connine et al. (1991) showed that listeners often relied on the biasing word at the end of the sentence to disambiguate the target word, when the target word had an ambiguous onset, whereas they were not biased by the context (biasing word), when the target word had a phonetically clear onset. They concluded that top-down inference from the context is given more weight when the target input is ambiguous than when it is clear.

In the same study, Connine et al. (1991) showed that the contextual biasing effect was present when the target word was separated from the disambiguating word by a small number of syllables (NEAR condition) but not when there was a larger number of syllables (FAR condition). The response time data, however, showed that in the FAR

condition, most of the time, the decision was being made prior to the availability of the disambiguating information, suggesting that there was an approximately 1 s window to make a decision based on acoustic-phonetic information prior to its decay.

In an eye-tracking study, Brown-Schmidt & Toscano (2017) showed a context bias effect even when the ambiguous word is separated from the biasing context by six-seven syllables. In fact, prior to disambiguation, the listeners fixated on the interpretation of the word that did not match the context but shifted their gaze only after having heard the biasing context. Similarly, Szostak and Pitt (2013) replicated the contextual biasing effects on ambiguous-sounding phoneme identification. Although smaller than in the NEAR condition, they also observed a biasing effect in the FAR condition. The authors suggested that the temporal window for disambiguating unclear acoustic-phonetic information may not be completely fixed, as suggested by Connine et al. (1991), but rather influenced by other factors, such as syntactic complexity or experience with language use.

Another factor that could affect the temporal window may be the typological characteristics of a given language. However, so far, language processing studies have mainly focused on English, therefore making it difficult to generalize the findings to other languages. In fact, it is debated whether all languages are processed in the same way and thus findings in one are generalizable to the others (Pinker, 1994), or whether each language has its unique characteristics, shaped by language users (Evans & Levinson, 2009). In the current study we address the question of whether individual languages are all processed in the same way or afford different processing strategies. Specifically, we investigate potential differences in the processing of the two languages—Danish and Norwegian—which are closely related but differ substantially in their phonological structure.

The Case of Danish and Norwegian

The relative weight that context is given in speech comprehension may vary from language to language, depending on the typological characteristics of a given language. We hypothesized that Danish may be a language, where top-down contextual processes is given larger weight than bottom-up acoustic-phonetic cues, compared to its close linguistic neighbors, Swedish and Norwegian. In terms of cross-linguistic comparisons, Danish and Norwegian thus allow for a well-controlled natural experiment. Denmark and Norway have a long common history, and have strong similarities in culture, education, politics, and other extra-linguistic factors. The two languages also have very similar grammars, morphology, and vocabulary—but differ in their phonology: Danish has a much more opaque phonology than Norwegian.

The sound structure of Danish is quite unique. Apart from having an unusually high number of vowels and vowel-like consonants, there is also a higher degree of syllabic reduction and assimilation of both vowels and consonants, compared to its close relatives Norwegian and Swedish (Basbøll, 2005). As a result, Danish is more difficult to acquire as a native language than Swedish and Norwegian (Bleses, Basbøll & Vach, 2011). There is also evidence that out of these three mutually intelligible Scandinavian languages, Danish is the most difficult to understand (Gooskens et al., 2010; Hilton, Schüppert & Gooskens, 2011). This may be due to the fact that there is generally a higher degree of syllabic reduction in Danish than in Norwegian and Swedish. Moreover, due to phonological reduction in Danish, some words sound identical to each other (Basbøll, 2005). In general, Danish speakers are thus exposed to a more imperfect and unclear acoustic-phonetic input compared to their Scandinavian neighbors. And, as a result, Danish speakers may rely on top-down processes to a larger extent than Norwegian and Swedish speakers.

In the current study, we adapted the paradigms used by Connine et al. (1991) and Szostak and Pitt (2013) to test the hypothesis that Danish speakers, due to the phonological peculiarities of the language, rely more on top-down processes than Norwegian speakers do. We predicted that when presented with ambiguous sounding words, Danish speakers would rely more on contextual cues compared to Norwegian speakers. In fact, for Danish speakers, we expected this effect to be present not only in the NEAR condition but also in the FAR condition, indicating that the acoustic-phonetic bottom-up input is given relatively less weight by Danish speakers than by Norwegian speakers. Moreover, we predicted that Danish speakers would be more inclined to wait until the end of the sentence to respond than Norwegian speakers (H1: language main effect). Following the findings for English by Szostak and Pitt (2013) and Connine et al. (1991), we expected that both Danish and Norwegian speakers would be affected by contextual bias (H2: contextual bias main effect) and that the effect would be stronger in the NEAR condition (H3: bias by distance interaction). Additionally, given the processing differences between Danish and Norwegian, we expected the bias effect to be stronger in Danish (H4: bias by language interaction) and the bias by distance interaction stronger in Norwegian (H5: bias by distance by language interaction).

To test these hypotheses, we fitted our experimental data to a drift diffusion model (Ratcliff, 1978), which jointly takes into account responses and response times as dependent variables and allowed us to separate the time preceding the decision making process (non-decision time), the rate at which evidence is accumulated (drift rate) and the amount of evidence needed to make a decision (boundary separation, see Methods for details). We expected to observe a longer non-decision time in Danish speakers than Norwegian speakers (H1). We expected the evidence accumulation rate to be affected by contextual bias (faster in congruent contexts, H2). We expected both drift rate and boundary separation to be affected by contextual bias in a way that is modulated by distance (stronger effect for the shorter distance, H3), and by language (Danish speakers being more sensitive to context, H4). Finally, we expected both drift rate and boundary separation to follow H5: Norwegian speakers

will show a stronger bias by distance interaction, that is, the way distance modulates contextual bias will be more marked for them (H5).

Method

Participants

Thirty-two Danish (22 female, age = 19 - 36 years, median = 23, sd = 3.3) and 34 Norwegian (13 female, age = 19 - 28 years, median = 22, sd = 2.5) right-handed native speakers participated in the study. The participants did not report a history of neurological or psychiatric disorders. The Danish speakers were tested at the Cognitive and Behavior Lab at Aarhus University in Denmark, while the Norwegian speakers were tested at the Faculty of Humanities at the University of Bergen in Norway. All participants received a monetary compensation for their participation.

Materials

We constructed 16 pairs of carrier sentences, half of which were biased towards the target word *sendt,* as shown in (1a) (Danish) and (1b) (Norwegian) and the other half towards *tændt* in Danish (2a) or *tent* in Norwegian (2b). In 8 pairs, the distance between the target and the disambiguating word was one syllable (NEAR condition); and in the remaining 8 pairs, it was 5-7 syllables (FAR condition). Importantly, in normal speech, except for the difference in the initial phoneme, the two target words have similar (rhyme) endings in both languages.

(1a) *Hun har **sendt** en (imponerende klar) mail.*
[ˈhun ˈhɑ ˈ**sɛntˢ** eˀn (empoˈneˀʌnə klɑˀ) ˈmɛjl]
'She has **sent** an (impressively clear) email.'
(1b) *Hun har **sendt** en (imponerende klar) mail.*
[ˈhʉn ˈhɑr ˈ**sɛnt** en (empoˈnerʌnə klɑr) ˈmɛjl]
(2a) *Hun har **tændt** en (imponerende klar) lampe.*
[ˈhun ˈhɑ ˈ**tˢɛntˢ** eˀn (empoˈneˀʌnə ˈklɑˀ) ˈlɑmbə]
(2b) *Hun har **tent** en (imponerende klar) lampe.*
[ˈhʉn ˈhɑr ˈ**tɛnt** en (empoˈnerʌnə klɑr) ˈlɑmpə]
'She has **turned-on** a(n) (impressively clear) lamp.'

Both the Danish and the Norwegian stimuli were recorded by a native male speaker of the respective languages. The recorded Danish [s] and [t^s] sounds in the target words *sendt* and *tændt* differed primarily according to the duration of the frication noise, the rise time of the noise, and the duration of the silent interval between noise offset and onset of the following vowel. The same was true for the Norwegian target words' [s] and [t^h] sounds, which in addition differed in intensity. A ten-step s-t continuum was generated for each language by interpolating between the endpoints according to the above-mentioned acoustic differences and splicing the resultant sounds to a single token of *tændt/tent*. The continua had a clear [s] at one end and a clear [t^s] (Danish) or [t^h] (Norwegian) at the other end and with eight intermediate steps.

We then piloted the two continua (forced choice identification). Based on the identification functions we chose steps 4, 5 and 6 as they straddled the mean category boundaries in each language. These three intermediate steps and the endpoints were used in the experiment. Thus, there were 160 trial sentences in total. The experiment was programmed and carried out in PsychoPy2 v1.90.1. (Peirce & MacAskill, 2018).

Procedure

Prior to the experiment, the participants received detailed instructions on the screen in their native languages. They were told to indicate which word they thought they heard and they were warned that sometimes this would not be easy. The participants were also instructed that they could use any information in the sentence that may help them to make their decision (cf. Connine et al., 1991; Szostak & Pitt, 2013). Following the instructions, the participants completed a practice trial and then the real experiment began. The target words *sendt* and *tændt/tent* were presented in boxes in the upper left and right corners of the screen while the target sentences were played back through headphones. The participants responded by clicking on the appropriate word with the mouse. They were allowed to respond at any point during and after the sentence playback (cf. Connine et al., 1991). There was a pause of 1.5 s between each trial, during which a blank screen was presented. The 160 stimuli were presented in a pseudorandomized order across four blocks of 40 trials. The first two items of the experiment contained the endpoints [s] and [t^s]/[t^h], respectively, in a congruent context. After each block, the participants had a self-paced short break. The whole experiment took 15 – 20 minutes. Responses and response times (RTs) were recorded as dependent variables. RTs were measured from the onset of the target word until the mouse click.

Data Analysis

Mouse clicks outside the boxes were recorded as missing values and were removed from the analysis. Further, responses corresponding to RTs higher than 3 standard deviations from the mean (> 8s) were also excluded from the analysis (2% of the total number of data points).

We fitted a Bayesian multilevel drift diffusion model (DDM) to the response and RT data. DDM is a sequential sampling model that explains cognitive processes underlying decision-making in 2-choice discrimination tasks (Ratcliff & McKoon, 2008). Decisions are described by the following parameters: the drift rate (δ) is the average rate of evidence accumulation; the boundary separation (α) is the evidence necessary to make a decision; the starting point (β) is the initial bias towards one of the response boundaries; and non-decision time (τ) is the part of the response time that is not involved in evidence accumulation (e.g., motor response execution). We conditioned drift rate and boundary separation on language, contextual bias (congruent/ incongruent), distance (NEAR/FAR) and continuum step as fixed effects, including their interactions, and participants as

varying effects, including varying slopes for bias, distance and step. We assumed no biased preference for a specific response and conditioned non-decision time on language and contextual bias only due to convergence issues. PSIS-LOO model comparison was used to select the relevant predictors to include (Vehtari, Gelman & Gabry, 2017), which led us to exclude step. We set weakly informative priors for δ (mean = 0, *sd* = 0.5), α (mean = 1.5, *sd* = 1) and τ (mean = 0.2, *sd* = 0.1). Model quality was thoroughly assessed via predictive prior and posterior checks, Rhat and divergence diagnostics. The model presented no divergences, and all chains mixed well and produced comparable estimates (Rhat < 1.01). In order to assess the evidence in favor or against our hypotheses, we used Evidence Ratio (ER, a generalization of Bayes factors allowing for directional hypotheses). An ER above 3 indicates moderate to substantial evidence for our hypothesis, below 0.3 indicates moderate to substantial evidence for the null hypothesis, and anything in between is inconclusive evidence (Morey, Rouder & Jamil, 2014). The models were implemented through the *brms* (Bürkner, 2017) and *RWiener* (Wabersich & Vandekerckhove, 2014) packages in RStudio v1.1.46, following the procedures of the tutorial written by Singmann (2017).

Results

Descriptive statistics are presented in Table 1. Full parameter estimates by condition are presented in Table 2.

Table 1: Mean reaction times ± standard deviations (in seconds) and *tændt/tent* response mean proportions ± standard deviations for Danish and Norwegian and NEAR and FAR distances with the context biased towards *sendt* or *tændt/tent*.

Language	Distance	Context bias	RT (s)	Response *tændt/ tent* (%)
Danish	NEAR	sendt	2.08 ± 0.82	66 ± 12
		tændt	2.15 ± 0.90	72 ± 8
	FAR	sendt	2.70 ± 1.07	66 ± 12
		tændt	2.71 ± 1.10	69 ± 10
Norwegian	NEAR	sendt	2.49 ± 0.98	34 ± 16
		tent	2.56 ± 1.02	50 ± 24
	FAR	sendt	3.38 ± 1.26	32 ± 18
		tent	3.35 ± 1.22	48 ± 22

As predicted by H1, we observed substantial evidence for non-decision time being longer in Danish than in Norwegian in congruent context ($\Delta\tau = 0.11 \pm 0.02$, ER > 1000), indicating that Danish speakers waited longer before starting to make a decision.

As per H2, we found substantial evidence for contextual bias affecting Danish speakers in the NEAR condition. When the response (i.e., *tændt*) matched the contextual bias (biased towards *tændt*, congruent context), evidence accumulation was faster ($\Delta\delta = 0.21 \pm 0.1$, ER = 45.5), than when the context did not match (biased towards *sendt,* incongruent context). There was also evidence that the boundary separation was larger for congruent context than for incongruent context ($\Delta\alpha = 0.98 \pm 0.29$, ER > 1000). Contrary to our expectations, however, there was no evidence that Norwegian speakers were affected by contextual bias ($\Delta\delta$ = -0.06 ± 0.12, ER = 0.46; $\Delta\alpha$ = -0.2 ± 0.36, ER = 0.41).

As expected (H4), we found substantial evidence for the bias effect being larger for Danish than for Norwegian speakers ($\Delta\Delta\delta = 0.27 \pm 0.12$, ER = 89.9, $\Delta\Delta\alpha = 1.18 \pm 0.39$, ER = 999). In other words, Danish speakers relied more on contextual evidence: matching context sped up their evidence accumulation more than for Norwegian speakers.

Table 2: The estimates of the diffusion drift model parameters per condition Bias (congruent/incongruent), Language (Danish/Norwegian) and Distance (NEAR/FAR). The parameters are drift rate (δ), boundary separation (α) and non-decision time (τ).

δ	estimate	95% CI
congruent:Danish:NEAR	2.12	1.94 - 2.31
incongruent:Danish:NEAR	1.91	1.75 - 2.07
congruent:Norwegian:NEAR	1.92	1.72 - 2.12
incongruent:Norwegian:NEAR	1.98	1.80 - 2.16
congruent:Danish:FAR	1.79	1.59 - 1.98
incongruent:Danish:FAR	1.74	1.56 - 1.93
congruent:Norwegian:FAR	1.70	1.47 - 1.92
incongruent:Norwegian:FAR	1.63	1.43 - 1.84
α		
congruent:Danish:NEAR	3.90	3.00 - 4.72
incongruent:Danish:NEAR	2.92	2.13 - 3.64
congruent:Norwegian:NEAR	4.31	3.27 - 5.29
incongruent:Norwegian:NEAR	4.52	3.58 - 5.41
congruent:Danish:FAR	3.25	1.94 - 4.48
incongruent:Danish:FAR	2.63	1.36 - 3.86
congruent:Norwegian:FAR	4.09	2.69 - 5.47
incongruent:Norwegian:FAR	3.55	2.16 - 4.98
τ		
congruent:Danish	0.57	0.54 - 0.6
incongruent:Danish	0.71	0.68 - 0.73
congruent:Norwegian	0.46	0.43 - 0.49
incongruent:Norwegian	0.50	0.46 - 0.54

We found moderate evidence that the drift rate was affected by contextual bias more in the NEAR condition than in the FAR condition in Danish speakers (H3: bias by distance interaction, $\Delta\Delta\delta = 0.15 \pm 0.17$, ER = 6.1). There was, however, no substantial evidence for boundary separation being affected by contextual bias differently according to distance ($\Delta\Delta\alpha = 0.37 \pm 0.58$, ER = 2.9). As for Norwegian speakers, there was no evidence either for drift rate ($\Delta\Delta\delta$ = -0.12 ± 0.17, ER = 0.3) or boundary separation ($\Delta\Delta\alpha$ = -0.75 ± 0.68, ER = 0.16) being affected more in the NEAR than in the FAR condition (against H3). Finally, as predicted, distance did not affect Norwegian speakers as much as

Danish speakers (H5, $\Delta\Delta\Delta\delta$ = -0.29 ± 0.18, ER = 16.7; $\Delta\Delta\Delta\alpha$ = -1.11 ± 0.71, ER = 15.7). This is likely to be due to the absence of bias effect in Norwegian altogether. The DDM simulations per each condition are depicted in Figure 1.

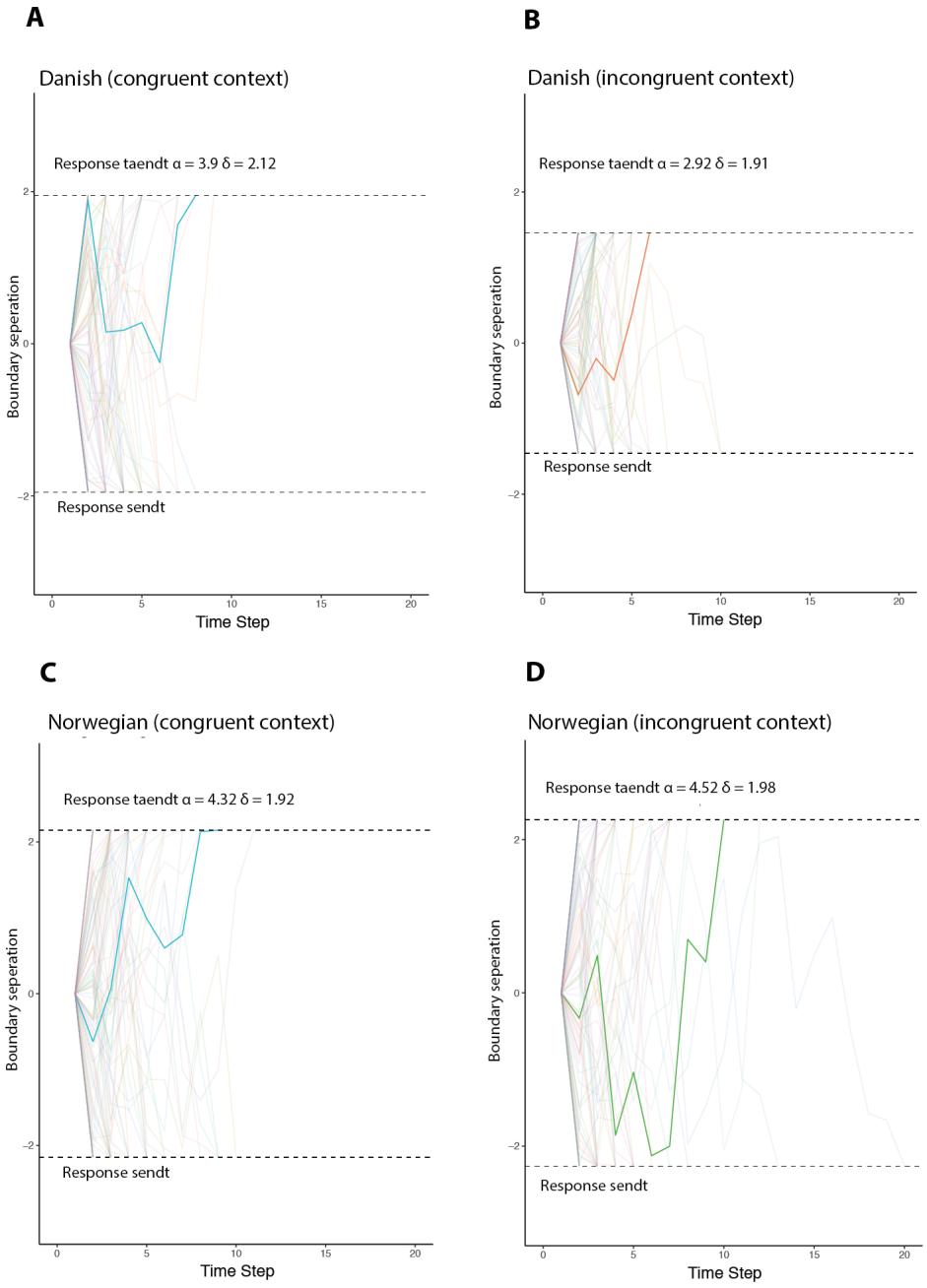

Figure 1: Simulation of the drift diffusion model for distance NEAR and congruent context (i.e. *tændt* bias) in A) Danish and C) Norwegian, and incongruent context (i.e. *sendt* bias) in B) Danish and D) Norwegian. The upper decision boundary is for the response *tændt* and the lower decision boundary is for the response *sendt*. The distance between the two boundaries is the boundary separation (α) and the evidence accumulation speed is the drift rate (δ). While there is no credible difference in drift rate between congruent and incongruent contexts for Norwegian, there is evidence that the drift rate is smaller in the incongruent context than in the congruent context in Danish. Thus, when the context is incongruent, evidence is accumulated slower to make a decision about *tændt*. The highlighted line is an example of the decision process in each condition.

Discussion

In the current study, we investigated whether contextual bias has a different effect on word recognition across the two related languages, Danish and Norwegian, and whether the distance between the target word and the disambiguating word affected word recognition across the two languages. We fitted our data to a drift diffusion model to obtain more subtle evidence about the cognitive processes underlying word recognition.

We found strong evidence that contextual bias affected the drift rate (the speed with which evidence is accumulated) in the NEAR condition in Danish. This indicates that acoustic-phonetic information alone is insufficient to make a decision and thus that additional evidence, such as contextual cues are integrated to support top-down processes of word comprehension. These findings are in line with previous evidence by Szostak and Pitt (2013) as well as Connine et al. (1991).

Surprisingly, we did not find evidence for contextual bias effects in Norwegian, which contradicts the previous evidence for English (Brown-Schmidt & Toscano, 2017; Bushong & Jaeger, 2017; Connine et al., 1991; McMurray et al., 2009; Szostak & Pitt, 2013). It is possible that this is because top-down contextual information is assigned even less weight in speech processing in Norwegian than in English or Danish. Moreover, Norwegian speakers may have responded prior to hearing the biasing context, thus not having the opportunity of using contextual information. In line with this, we found that Danish speakers generally wait longer to respond than Norwegian speakers (longer non-decision time in Danish compared to Norwegian, H1), which may be additional evidence that Danish speakers weight top-down contextual information more than bottom-up acoustic-phonetic information compared to Norwegians.

In line with our hypotheses, we found that Danish speakers were more affected by contextual biases than Norwegian speakers (H4), and that the contextual bias was stronger for Danish speakers in the NEAR condition than in the FAR condition, compared to Norwegian speakers (H5). However, importantly, the H3 interaction results held only for Danish, likely due to the lack of a contextual bias effect in Norwegian.

There was also some evidence that the bias effect on drift rate was stronger in the NEAR condition than in the FAR condition for Danish speakers but there was no credible evidence of the same effect on boundary separation. It is possible that a similar amount of information is necessary to make a decision about an ambiguous target word, as the nature of the information does not change across NEAR and FAR distances (i.e., the acoustic-phonetic cues are equally ambiguous and the disambiguating words remain the same). However, the speed at which this information is accumulated changes slightly. As Szostak and Pitt (2013) and Connine et al. (1991) suggested, there is a short temporal window to make a decision about the acoustic-phonetic information. Thus, in the NEAR condition, due to a higher drift rate in Danish speakers, it takes shorter time to choose a response that is congruent with the contextual bias (i.e., to respond *tændt* in a *tændt*-biased context).

The above-mentioned effect on drift rate was stronger for Danish than for Norwegian, indicating that the temporal window suggested by Szostak and Pitt (2013) may indeed vary due to different factors, in this particular case, phonological differences between languages. We interpret this evidence as suggesting that top-down contextual inferences are more important for Danish speakers compared to Norwegian speakers, when faced with acoustic-phonetically ambiguous stimulus. This may be because of the unique sound structure of Danish, which results in relatively more ambiguity in Danish speech than in other Scandinavian languages (Basbøll, 2005; Hilton et al., 2011; Gooskens et al., 2010). Thus, in line with first language acquisition studies (Bleses et al., 2008; 2011), we provide evidence that Danish is processed differently also by adult native speakers, compared to native Norwegian speakers.

It is possible that allowing participants to respond at any time during a trial may also have affected our results. Using the Connine et al. (1991) paradigm, Bushong & Jaeger (2017) showed that the context effect was smaller in the FAR condition, when the listeners could respond whenever they wanted. However, there was no difference between the NEAR and FAR conditions, when the listeners were forced to wait until hearing the biasing word to respond. In fact, the observation that participants change their response profile when forced to wait to the sentence offset, as shown by Brown-Schmidt & Toscano (2017), indicates that indeed free and forced responses may influence the decisions that listeners make. Thus, a future study comparing forced and free responses may shed light on the different strategies Danish and Norwegian speakers may be using when completing the task.

The current study, however, has one important limitation: the steps of the [s]-[t^s]/[t^h] continuum were not included in the DDM model. Step is a crucial feature and it could provide more nuanced information not only about the contextual bias and distance effect on word recognition processes but also how these processes vary cross-linguistically. Future work should include a nuanced modeling of step (e.g., as a monotonic but not necessarily a linear function) to assess whether step can be meaningfully included and help better explain the data. We anticipate that such analyses might provide a more detailed picture of the points in the continuum at which information is accumulated faster and at which more information is needed. Thus, a more complex drift diffusion model with the steps of the continuum as one of the fixed effect variables would shed further light on the cognitive processes underlying spoken word recognition when the acoustic-phonetic cues are ambiguous.

Despite these limitations, our study suggests that Danish is processed differently compared to Norwegian. When exposed to ambiguous stimuli, Danish speakers rely more on top-down processes than Norwegian speakers. Contrary to the standard view that all languages are equally easy to learn and use (e.g., Pinker, 1994), we provide evidence that

languages can differ in how they are processed, as suggested, for instance, by Evans and Levinson (2009)— and that there may be a continuum of reliance on top-down processes, where English could be lying somewhere between Danish and Norwegian. However, future cross-linguistic studies are necessary to confirm this assumption.

Acknowledgements

This study was supported by Danish Council for Independent Research (FKK) Grant DFF-7013-00074 awarded to Morten H. Christiansen.

References

Basbøll, H. (2005). *The phonology of Danish*. Oxford University Press.

Bleses, D., Basbøll, H., & Vach, W. (2011). Is Danish difficult to acquire? Evidence from Nordic past-tense studies. *Language and Cognitive Processes*, 26(8), 1193-1231.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language*, 35(3), 619-650.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat:" The effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, 103(5), 2670-2676.

Brown-Schmidt, S., & Toscano, J. C. (2017). Gradient acoustic information induces long-lasting referential uncertainty in short discourses. *Language, Cognition and Neuroscience*, 32(10), 1211-1228.

Bushong, W., & Jaeger, T. F. (2017). Maintenance of Perceptual Information in Speech Perception. *In CogSci*.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1-28.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, 30(1), 234.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32, 429-448.

Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language*, 44(3), 325-349.

Gooskens, C., Van Heuven, V. J., Van Bezooijen, R. & Pacilly, J. J. (2010). Is spoken Danish less intelligible than Swedish? *Speech Communication*, 52, 1022-1037.

Hilton, N. H., Schüppert, A., & Gooskens, C. (2011). Syllable reduction and articulation rates in Danish, Norwegian and Swedish. *Nordic Journal of Linguistics*, 34(2), 215-237.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10(1), 29-63.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of memory and language*, 60(1), 65-91.

Morey, R. D., Rouder, J. N., & Jamil, T. (2014). BayesFactor: Computation of Bayes factors for common designs (Version 0.9. 9).

Peirce, J. W., & MacAskill, M. R. (2018). Building Experiments in PsychoPy. London: Sage.

Pinker, S. (1994). *The language instinct*. New York: William Morrow & Co.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, 20(4), 873-922.

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, 110(4), 474.

Singmann, H. (2017, November 26). Diffusion/Wiener Model Analysis with brms – Part I: Introduction and Estimation [Blog post]. Retrieved from http://singmann.org/wiener-model-analysis-with-brms-part-i/

Szostak, C. M., & Pitt, M. A. (2013). The prolonged influence of subsequent context on spoken word recognition. *Attention, Perception, & Psychophysics*, 75(7), 1533-1546.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413-1432.

Wabersich, D., & Vandekerckhove, J. (2014). The RWiener package: An R package providing distribution functions for the Wiener diffusion model. *The R Journal*, 6(1), 49–56.