Subtitles supplemental materials

Experiment 3

Figure 1.

Subject's ratings about how well they would be able to follow instructions in Danish by condition (subtitles, no subtitles).

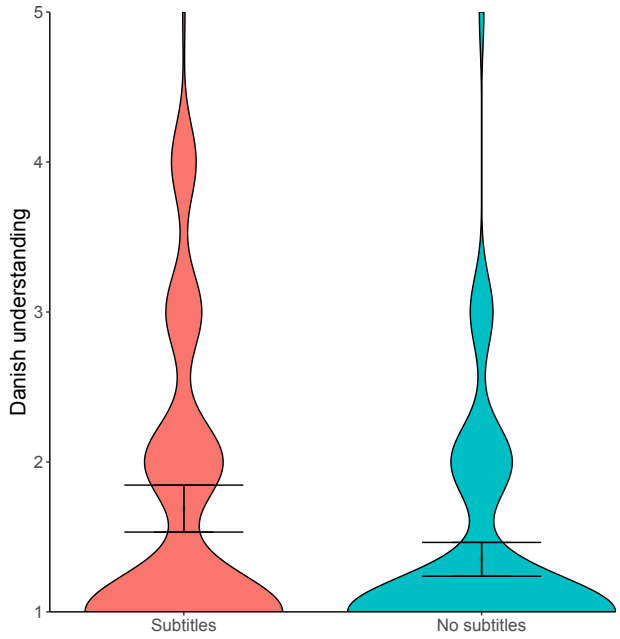


Figure 2.

Subject's ratings about how well they would be able to understand the Danish TV news by condition (subtitles, no subtitles).

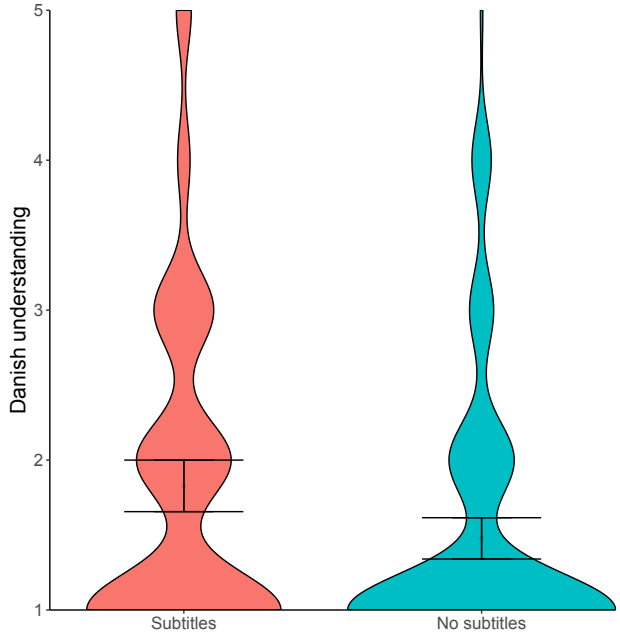


Figure 3.

Subject's ratings about how well they would be able to understand the Danish weather forecast by condition (subtitles, no subtitles).

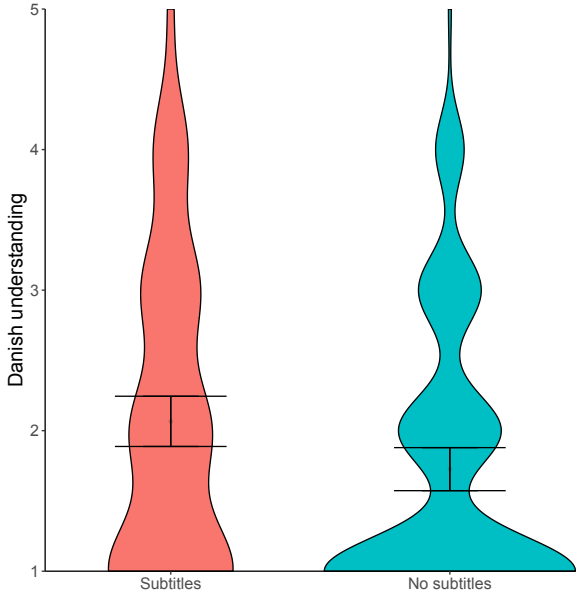


Figure 4.

Subject's ratings about how well they would be able to make friends with peers who speak only Danish by condition (subtitles, no subtitles).

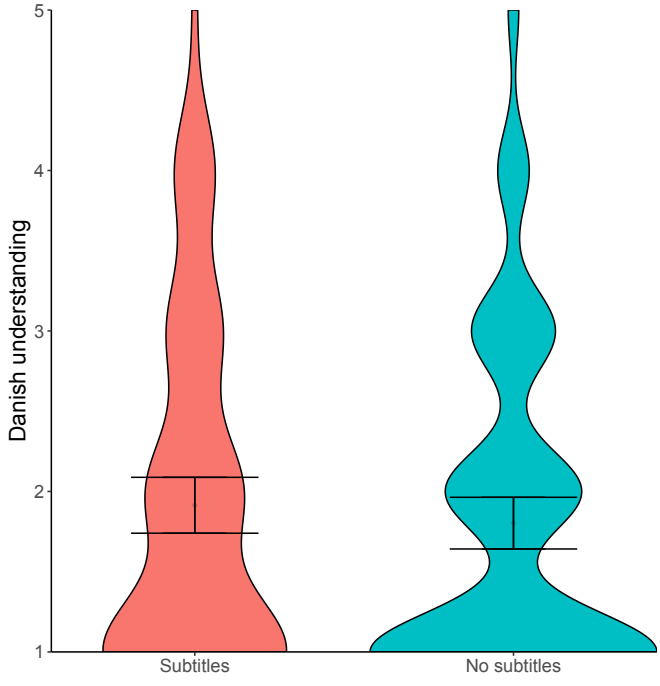


Table 1.

Correlations between four dependent measures (Instructions, News, Weather, & Friends).

	Instructions		News		Weather		Friends	
	r	95% CI	*r*	95% CI	*r*	95% CI	*r*	95% CI
Instructions	1.00		.57	[.49, .64]	.51	[.42, .59]	.45	[.35, .53]
News	.57	[.49, .64]	1.00		.57	[.49, .64]	.39	[.29, .48]
Weather	.51	[.42, .59]	.57	[.49, .64]	1.00		.38	[.28, .47]
Friends	.45	[.35, .53]	.39	[.29, .48]	.38	[.28, .47]	1.00	

Experiment 3 follow-up

As part of routine analysis to assess new materials, we received feedback from colleagues that they found the new "Borgen" debate harder to follow than the "Rita" school tour. On the basis of this feedback, we examined our data to determine if the two different videos produced different ratings.

Method

Subjects

We collected data from 138 MTurk workers. We analysed the data from 137 subjects after excluding 1 person who spoke Danish (M_{age} = 40.34, SD_{age} = 14.38; 15% identified as men, 82% identified as women, 3% gender diverse; 91% English first language).

Design

We used a 2(subtitles: subtitles, no subtitles) x 2(clip type: "Rita", "Borgen") mixed design. Subtitles was a between-subjects factor and clip-type was a within-subjects factor.

Procedure

First, subjects were randomly assigned to watch one of the two videos from Experiment 3, either with subtitles or without subtitles. One video was the "Rita" school tour; the other video was the "Borgen" chaotic political debate. Then subjects answered 7

questions about how easy it was to understand and follow the storyline: “How clear did you find the story in the film?” (*1 - Not at all, 7 - Completely*), “How cohesive did you find the story in the film?” (*1 - Not at all, 7 - Completely*), “How coherent did you find the story in the film?” (*1 - Not at all, 7 - Completely*), “How well do you believe you understood what happened in the story?” (*1 - Not at all, 7 - Completely*), “Did the story in the film make sense to you?” (*1 - Not at all, 7 - Completely*), “As you watched the film, how easy was it for you to predict what would happen next?” (*1 - Not at all easy, 7 - Very easy*), “Did the sequence of events in the film have a logical flow?” (*1 - Not at all, 7 - Completely*). Then, subjects watched the other video (“Rita” or “Borgen”) and made all the same ratings.

Results

Table 2.

Subjects’ mean ratings about how easy it was to understand and follow the storyline of the Rita and Borgen clips

	“Rita” clip						“Borgen” clip					
	Subtitles			No Subtitles			Subtitles			No Subtitles		
Rating	*Mdn*	M	95% CI	*Mdn*	*M*	95% CI	*Mdn*	*M*	95% CI	*Mdn*	*M*	95% CI
Clear	5.00	5.09	[4.36, 5.63]	4	3.62	[3.08,4.16]	4	3.48	[2.92, 4.05]	2	2.63	[2.20, 3.05]
Cohesive	5	4.97	[4.48, 5.46]	4	3.73	[3.28, 4.17]	4	3.70	[3.28, 4.12]	3	3.20	[2.62, 3.78]
Coherent	5	5.03	[4.52, 5.55]	4	3.68	[3.26, 4.09]	4	3.55	[3.08, 4.02]	2	2.37	[1.95, 2.79]
Understood what happened	5	5.25	[4.81, 5.69]	4	3.65	[3.09, 4.21]	4	3.73	[3.14, 4.32]	2	2.34	[1.95, 2.73]
Make sense	6	5.41	[4.87, 5.95]	4	3.59	[3.04, 4.14]	4	3.76	[3.15, 4.37]	2	2.31	[1.88, 2.75]
Predict	3.5	3.97	[3.47, 4.47]	3	3.27	[2.73, 3.81]	3	3.03	[3.47, 3.59]	2	2.17	[1.77, 2.58]
Logical flow	5	5.19	[4.75, 5.62]	5	5.14	[4.70, 5.57]	4	3.91	[3.41, 4.41]	4	3.91	[3.37, 4.46]

We report the results in two ways. First, because our primary interest was in how easy people would understand and follow the storyline in these two clips that would be we used in a between-subjects design, we compared the ratings for the clip subjects saw first. We carried out 2 (subtitles, no subtitles) x 2("Rita", "Borgen") analyses on subjects' ratings for each of the 7 items. As shown in the table above, compared to subjects who watched the "Borgen" clip, subjects who watched the "Rita" clip found the video clearer ($F(1, 136) = 25.98$, $p < .001$), more cohesive ($F(1, 136) = 14.02$, $p < .001$), more coherent ($F(1, 136) = 39.20$, $p < .001$), they better understood what happened ($F(1, 136) = 32.34$, $p < .001$), the film made more sense ($F(1, 136) = 30.80$, $p < .001$) and had more of a logical flow ($F(1, 136) = 27.63$, $p < .001$) and it was easier to predict what would happen next ($F(1, 136) = 16.82$, $p < .001$). Compared to subjects who watched the clip without subtitles, subjects who watched the clip with subtitles found the video clearer ($F(1, 136) = 20.81$, $p < .001$), more cohesive ($F(1, 136) = 13.02$, $p < .001$), more coherent ($F(1, 136) = 32.23$, $p < .001$), they better understood what happened ($F(1, 136) = 36.03$, $p < .001$), the film made more sense ($F(1, 136) = 38.03$, $p < .001$) and it was easier to predict what would happen next ($F(1, 136) = 9.83$, $p = .01$). They did not report that the clip had more of a logical flow ($F(1, 136) = .01$, $p = .92$).

Second, to determine the relative ease with which people understood and followed the storyline, we compared, within-subjects, the mean understanding rating for both clips. Compared to subjects who watched the clip without subtitles, subjects who watched the clip with subtitles found it easier to understand and follow the storyline ($F(1, 136) = 36.36$, $p = < .001$). Compared to subjects who watched the "Borgen" clip, subjects who watched the "Rita" clip found it easier to understand and follow the storyline ($F(1, 136) = 155.71$, $p = < .001$). There was an interaction between mean understanding for each clip and which clip subjects saw first ($F(1, 136) = 22.79$, $p = < .001$). Tukey's HSD tests indicated that subjects

found it easier to follow and understand the "Rita" clip if they watched it after the "Borgen" clip relative to subjects who watched the "Rita" clip first, M_{diff} = .98 [.47, 1.50], $p < 0.001$. Whereas subjects reported it was similarly easy to follow and understand the "Borgen" clip regardless of whether they saw that clip first or second M_{diff} = .17 [-.35, .70], $p = 0.82$.

These results suggest that subtitles enhanced comprehension of both videos and add support to our colleague's feedback that the "Borgen" video was more difficult to follow than the "Rita" clip.

Experiment 4 materials norming

The purpose of this experiment was to ensure that people found it easier to comprehend a series of ambiguous passages when they were accompanied by a short title describing the topic. We were also interested in how many Mechanical Turk workers were familiar with these ambiguous passages.

Method

Subjects

We aimed to collect data from 50 MTurk workers. But because MTurk and Qualtrics interact in such a way that it is possible to overcollect a target sample size, we collected data from 52 subjects. We analysed the data from 44 subjects after exclusions (M_{age} = 41.02, SD_{age} = 15.22; 32% identified as men, 68% identified as women; 95% English first language).

Design

We used a between-subjects design with two conditions (title: title, no title).

Procedure

Subjects were randomly assigned to read four difficult-to-comprehend passages in English, either with subtitles or without subtitles. The passages were about driving a car, flying a kite, a trip to space, and washing clothes (Bransford & Johnson, 1972; Wiley & Rayner, 2000). After each passage, subjects answered the following questions. “How difficult was it to comprehend the passage?” (*1= very easy, 7 = very difficult*). “Have you ever come across this passage before?” (*yes, no*). “How confident are you that you have seen this passage before?” (*1= not at all confident, 7 = very confident*)

Results

First, we excluded the data from 8 subjects for providing non-sensical responses. Then to determine how familiar people were with the passages, we measured the proportion of subjects who were confident they had come across each passage before. We pre-registered that if the proportion of subjects who rated confidence >4 was higher than 10%, we would start the norming process again with different passages. But there were zero exclusions on the basis of the criteria.

Everyone who read the untitled passage about driving a car correctly guessed what it was about. Moreover, those who read the titled version of the passage rated their comprehension as 3.08 on 5-point scale (where lower scores indicate easier comprehension) compared to their untitled counterparts who rated it as 3.32. Because of this trivial difference was markedly less than the comparable difference in the other 3 passages (untitled kite: 4.37; titled kite: 3.56; untitled space: 4.74, titled space: 3.72; untitled clothes: 4.21, titled clothes: 3.16). Therefore, we dropped this passage from the set of materials and proceeded with a set of three passages.

Experiment 4 pilot study

In Experiment 4 we aimed to develop new materials that would minimize social information and eliminate sociogestural information. Therefore, we removed all visual information so subjects could hear only the foreign audio, and adapted work by Bransford and Johnson (1972) and Wiley and Rayner (2000). More specifically, we adapted a difficult-to-comprehend passage, see for example Table 1. Because this work is so well known, it is important to make clear the significant ways in which our use of these materials was different. For instance, when Bransford and Johnson attached the title “washing clothes” to the passage in the left column of Table 1, it provided semantic context and helped people scaffold incoming information relative to subjects who saw just the passage without a title. By contrast, we asked our subjects to listen to this passage translated to Danish, as the right column of Table 1 shows. What then should be the effect of asking some people to listen to this Danish passage and telling them it was about "washing clothes”? Whereas, for Bransford and Johnson subjects, knowing the passage was about washing clothes made it easier to comprehend, but for our subjects it could not. The only thing that knowledge would do for them is let them know the foreign words they are hearing are about washing clothes, as illustrated in Panel C.

And now consider what should be the effect of showing some people English language subtitles and not telling them what the passage was about. This situation essentially turns the passage into the “difficult-to-comprehend” control condition in Bransford and Johnson, in which people read a passage without being told what it was about—here equivalent to Panel B. On the basis of this thinking, we reasoned that only the combination of information about the topic plus subtitles would be equivalent to the experimental “easier-to-comprehend” condition in Bransford and Johnson, and produce easier comprehension. Therefore, in we addressed this question. To what extent would people who heard the Danish passage—

knowing the topic or seeing English subtitles—mistake their ease of comprehension as evidence they could understand that language in new situations.

Before answering this question, we first normed the materials. We found these data support the hypothesis that, of the four conditions in Table 1, only the "topic plus subtitles" condition was easy to comprehend.

Table 1.

Example audio transcripts in Experiment 4.

Panel A **Topic "Washing clothes"** The procedure is actually quite simple. First you arrange things into different groups depending on their makeup. Of course, one pile may be sufficient depending on how much there is. If you have to go somewhere else due to lack of facilities that is the next step. It is important not to overdo it. It is better to do too few things at once than too many. This may seem unimportant, but complications from doing too many can easily arise. A mistake can be expensive. The manipulation of the appropriate mechanisms should be obvious. At first the whole procedure may seem complicated, but soon it will become just another facet of life. It is hard to foresee any end to the need for this task.	**Panel C** **Topic "Washing clothes"** Proceduren er egentlig ret simpel. Først organiserer du tingene i forskellige grupper afhængigt af deres udseende. Én bunke kan naturligvis være tilstrækkeligt, afhængigt af hvor meget der er. Hvis du bliver nødt til at tage et andet sted hen på grund af manglede faciliteter, er det det næste skridt. Det er vigtigt ikke at overgøre det. Det er bedre at ordne få ting på én gang fremfor for mange. Det kan virke trivielt, men der kan nemt opstå problemer, hvis man ordner for mange. En fejl kan være dyr. Mekanismerne og hvordan de påvirkes burde være indlysende. Til at begynde med kan hele proceduren virke kompliceret, men det bliver hurtigt bare endnu en del afl livet. Det er svært at forestille sig, at behovet for at udføre denne opgave forsvinder.
Panel B The procedure is actually quite simple. […] It is hard to foresee any end to the need for this task.	**Panel D** Proceduren er egentlig ret simpel. […] Det er svært at forestille sig, at behovet for at udføre denne opgave forsvinder.

Method

Subjects

We collected data from 333 MTurk workers, on the basis of no prior work with which to estimate the size of the effect, we aimed for at least 75 subjects for each condition, anticipating exclusions. We analysed the data from 314 subjects (M_{age} = 41.71, SD_{age} = 14.41; 38% identified as men, 61% identified as women, 1% gender diverse; 93% English first language).

Design

We used a 2(subtitles:subtitles, no subtitles) X 2(title:title and picture, no title or picture) between-subjects design.

Procedure

The experiment proceeded in three phases. In the first phase, subjects listened to one of three audioclips of people speaking a difficult-to-comprehend passage in Danish. Half the subjects were randomly assigned to listen to this passage while seeing English subtitles; at the same time these subjects were evenly divided between also seeing a picture (for instance, of a washing machine) and title ("washing clothes") or no picture or title. The other half of subjects were randomly assigned to listen to this passage without seeing English subtitles; again, half were evenly divided between also seeing a picture (for instance, of a washing machine) and title ("washing clothes") or no picture or title.

In the second phase, subjects answered the following four questions in a randomised order: "How well would you be able to follow Danish instructions in an emergency?" (*1 = Quite well, 5 = Not at all well*), "If you were a new student at a Danish school and your peers spoke only Danish, how well would you be able to make friends?" (*1 = Quite well, 5 = Not at all well*), "How well would you be able to understand the top story on the Danish TV

news?" (*1 = Quite well, 5 = Not at all well*), and "How well would you be able to understand the weather forecast on Danish TV?" (*1 = Quite well, 5 = Not at all well*). Again, subjects were asked "How much expertise do you think is involved in understanding a foreign language" (*1 = No expertise, 5= A great deal of expertise*) and asked to rate "How difficult was it to comprehend the video?" (1-*Very easy*, 7-*Very difficult*).

In the third phase, subjects listened to a series of Danish words (some they had heard in the video and some that are commonly used) and were asked to translate these words into English. We used the Lexiteria database to select commonly used words within the top 200 (https://lexiteria.com/word_frequency/danish_word_frequency_list.html). According to their website, "The Lexiteria Danish Word Frequency List 2010 contains 355,828 words taken from a 25,838,905 word corpus based on edited web pages in 2010."

Results

First, we looked at the distributions of subjects' responses to the four main DVs. We noticed that the "Make Friends" item had different patterns and upon further analysis, was less correlated with the other 3 confidence ratings. We also received comments from subjects suggesting this "Make Friends" item was less reliant on foreign language understanding, such as: "There have been several occasions in my life when I have made friends with people who did not speak English, and whose language I did not speak". Therefore, we decided not to include the item in the mean confidence rating, and instead we carried out 2(subtitles, no subtitles) x2(title, no title) analyses on subjects' mean confidence ratings for the remaining 3 DVs: following Danish instructions, understanding the first story on the Danish TV news, and understanding the Danish weather forecast. We found that the context of the video mattered. That is, people who watched the video with a title were more confident than people who watched the untitled video, ($F(1, 313) = 5.18$, $p = .02$). But those saw the subtitled

videos were more confident than those who did not ($F(1, 313) = 6.57$, $p = .01$). There was no evidence of an interaction. Pairwise comparisons indicated that compared to subjects who watched the Danish video with no subtitles or title, subjects who watched the video with subtitles and a title reported they were better able to understand Danish more generally (Tukey's HSD effect size: $M_{diff} = .27$ [0.01, 0.52], $p = .04$). All other conditions did not significantly differ from one another.

Next, we looked for evidence to support our hypothesis that the title and subtitles boost confidence because they ease comprehension. We carried out 2(subtitles, no subtitles) x 2(title, no title) analyses on subjects' responses to "how difficult was it to comprehend the video?" (1-*Very easy*, 7-*Very difficult*). We found an interaction, such that titles only improved comprehension of the video when people had subtitles ($F(1, 313) = 7.74$, $p = .01$). Pairwise comparisons showed that compared to subjects who watched the Danish video with no subtitles or title, subjects who watched the video with subtitles and a title reported the video was easier to comprehend (Tukey's HSD effect size: $M_{diff} = 2.55$ [1.94, 3.17], $p < .01$).

Finally, we calculated subjects mean percentage score on the translation test to see if subjects who received both the title and subtitles learned any Danish. But their performance on the Danish translation test was no better than subjects in the other conditions (Danish translation test mean score: $M < 0.01\%$ in all conditions).

Experiment 4

Figure 1.

Subject's ratings about how well they would be able to follow instructions in Danish by condition (subtitles, no subtitles).

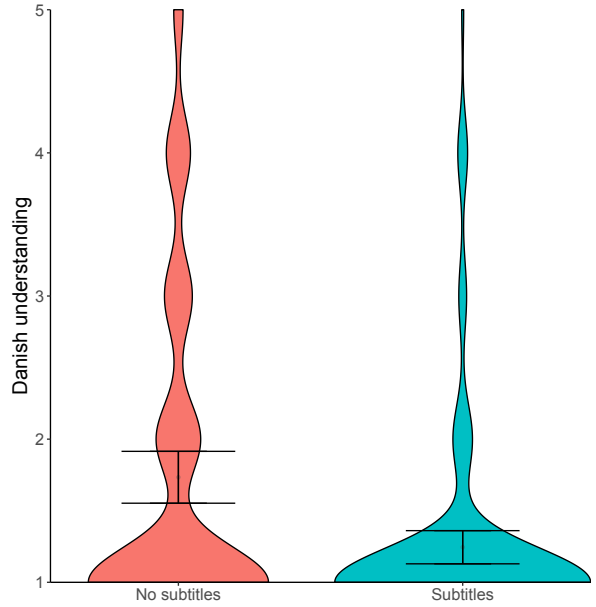


Figure 2.

Subject's ratings about how well they would be able to understand the Danish TV news by condition (subtitles, no subtitles).

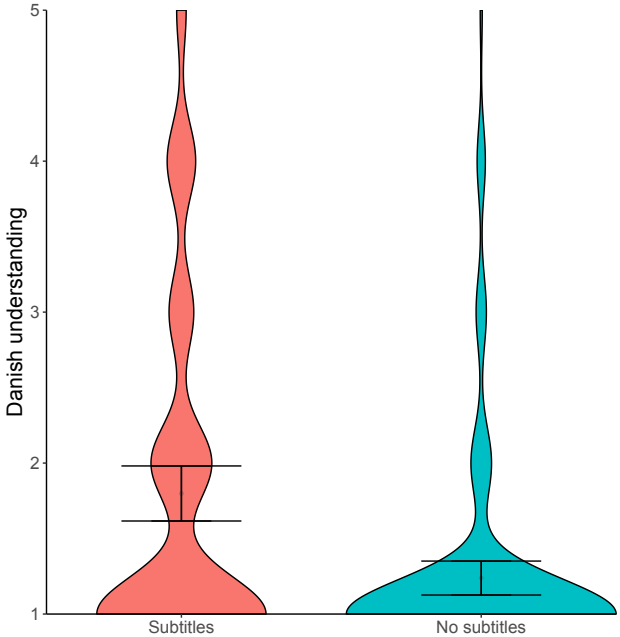


Figure 3.

Subject's ratings about how well they would be able to understand the Danish weather forecast by condition (subtitles, no subtitles).

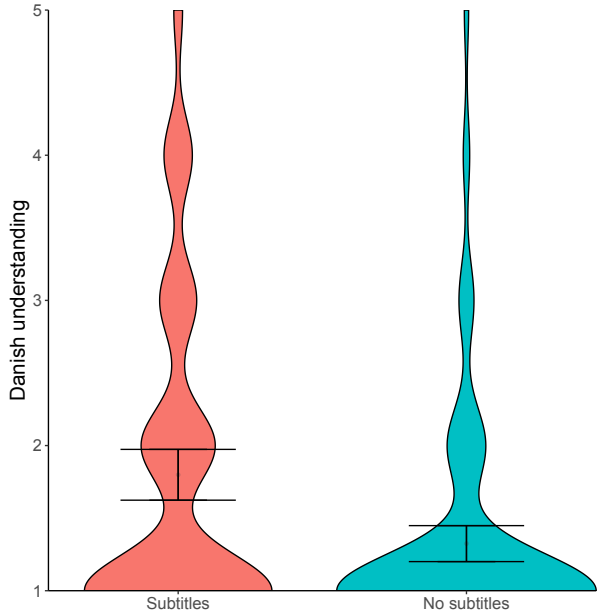
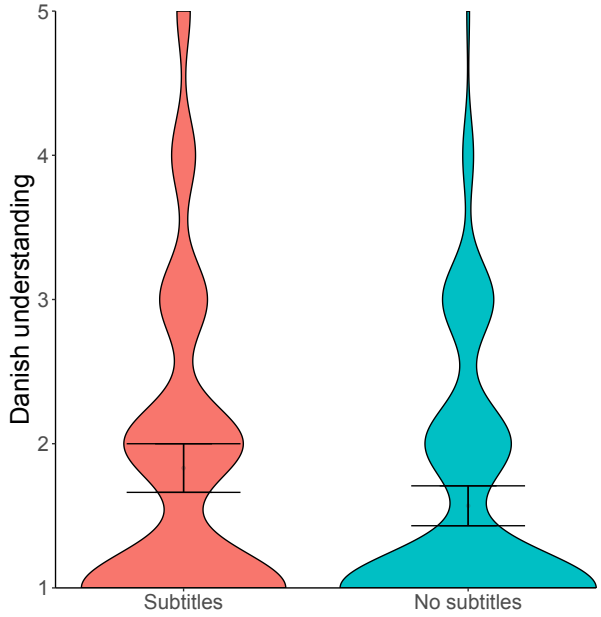


Figure 4.

Subject's ratings about how well they would be able to make friends with peers who speak only Danish by condition (subtitles, no subtitles).



Other pre-registered analyses

In our prior work (Jordan et al., 2022), we found men were more confident in their ability to perform a skill they had no prior experience in. In that study, we asked people to rate their confidence in their ability to land a plane and we were curious if we would observe a similar pattern when asking people about a different, but still highly specialized, skill. Therefore we preregistered between subjects two-way Anovas, and post-hoc tests, on subjects' language ability ratings with Subtitles(subtitles, no subtitles) and Gender as between subjects measures in Experiments 1a, 2 and 3.

We found no main effect of gender, or any interaction, in Experiment 1a and Experiment 3. However we did find a marginal main effect of gender in Experiment 2 ($F(1, 290) = 3.93$, $p = .05$). Because these findings did not replicate across experiments, we do not further speculate on any gender differences.

Compliance questions

Table 2.

Compliance questions used in experiments 1a, 1b, 2, 3 and 4.

Compliance question	Experiments used
On the previous page we told you whether the upcoming video would have subtitles. Will the upcoming video have subtitles? (yes, no)	1a, 1b
Could you hear the audio in the video? (yes, no)	1a, 1b
Did you watch the entire video, from start to finish? (yes, no)	1a, 1b, 2, 3
Did you watch the video more than once? (yes, no)	1a, 1b, 2, 3
Did you complete this session in a single sitting, without stopping? (yes, no)	1a, 1b, 2, 3, 4

Did you complete the experiment in an environment that is free of noise and distraction? (yes, no)	1a, 1b, 2, 3, 4

Additional information about videos

Table 3
Additional information about videos used in Experiments 1a, 1b, 2, and 3.

Video	Synopsis	Length (min:sec)	Source
Rita Danish video	A teacher shows a new hire around a school, and everyone is speaking Danish.	1:00	Television show *Rita* (Torpe & Kaalund, 2012)
Borgen Danish Video	Several politicians engaged in a live, chaotic, televised, debate.	1:00	Television show *Borgen* (Price et al., 2010)

References

Torpe, C. (Writer), Kaalund, L. (Director). (2012) Season 1, Episode 1. [Television series episode] In Leth, K., & Morthorst, J. (Producers), *Rita*. Denmark: SF Studios.

Price, A., Lindholm, T., Gjervig Gram, J. (Writers) & Kragh-Jacobsen, S. (Director). (2010) Season 1, Episode 1. [Television series episode] In Hammerich, C. (Producer), *Borgen*. Denmark: DR Fiktion.