End-of-Turn Detection

Sean Leishman

MInf Project (Part 1) Report
Master of Informatics
School of Informatics
University of Edinburgh

2023

Abstract

This skeleton demonstrates how to use the `infthesis` style for undergraduate dissertations in the School of Informatics. It also emphasises the page limit, and that you must not deviate from the required style. The file `skeleton.tex` generates this document and should be used as a starting point for your thesis. Replace this abstract text with a concise summary of your report.

Research Ethics Approval

Instructions: *Agree with your supervisor which statement you need to include. Then delete the statement that you are not using, and the instructions in italics.*
Either complete and include this statement:
This project obtained approval from the Informatics Research Ethics committee.
Ethics application number: ???
Date when approval was obtained: YYYY-MM-DD
[If the project required human participants, edit as appropriate, otherwise delete:]
The participants' information sheet and a consent form are included in the appendix.
Or include this statement:
This project was planned in accordance with the Informatics Research Ethics policy. It did not involve any aspects that required approval from the Informatics Research Ethics committee.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Sean Leishman*)

Acknowledgements

Any acknowledgements go here.

Table of Contents

Chapter 1

Introduction

The preliminary material of your report should contain:

- The title page.
- An abstract page.
- Declaration of ethics and own work.
- Optionally an acknowledgements page.
- The table of contents.

As in this example `skeleton.tex`, the above material should be included between:

```
\begin{preliminary}
    ...
\end{preliminary}
```

This style file uses roman numeral page numbers for the preliminary material.

The main content of the dissertation, starting with the first chapter, starts with page 1. ***The main content must not go beyond page 40.***

The report then contains a bibliography and any appendices, which may go beyond page 40. The appendices are only for any supporting material that's important to go on record. However, you cannot assume markers of dissertations will read them.

You may not change the dissertation format (e.g., reduce the font size, change the margins, or reduce the line spacing from the default single spacing). Be careful if you copy-paste packages into your document preamble from elsewhere. Some LaTeX packages, such as `fullpage` or `savetrees`, change the margins of your document. Do not include them!

Over-length or incorrectly-formatted dissertations will not be accepted and you would have to modify your dissertation and resubmit. You cannot assume we will check your submission before the final deadline and if it requires resubmission after the deadline to conform to the page and style requirements you will be subject to the usual late penalties based on your final submission time.

Chapter 2

Background Review

2.1 Turn-taking: From the Conversational Analaysis Perspective

Over the last few decades, psycolinguists have been fascinated with the complexity of the mechanisms of conversation along with the apparent ease with which speaker's are able to converse in a orderly and timely manner. Sacks et al. [1974] is a widely cited paper that outlines some general observations that has gone on to inform general turn-taking literature. They observed that turn-taking organisation is not planned in adavance however the actions taken are still coordinated, in a flexible manner that can be decided upon by the current participants in a conversation; typically one person speaks at a time and most transitions have a small gap or overlap but transitions do also occur with no gap and no overlap. Automatic analysis has managed to substantiate these claims with the existence of generally short turns (mean 1680ms, median 1227ms) (Levinson and Torreira [2015]) and that, the majority of turn transitions (51%-55%) take places under 200ms (Heldner and Edlund [2010]). An even greater majority of turn transitions take place between -100ms and 500ms (Levinson and Torreira [2015]). Turn-taking is finely tuned and managed.

2.1.1 Models of Turn-taking Organisation

Turn-taking organisation has generally been characterised in two different ways within literature: the `reactionary` and the `predictive` approach. The former assumes that participants simply understand end-of-turn signals and react to them accordingly while the predictive approach entails the listener predicting the end of turn in advance such that responses are well timed.

The reactionary approach assumes that turn-taking organisation is regulated by both vocal and gestural signals (Yngve [1970]). This approach was pioneered by (Duncan [1972, 1973, 1974], Duncan and Fiske [2015]) who argued for a precise set of context free turn-yielding 'signals'. Duncan [1972] described phrase-final intonation, drawl on the final syllable, termination of hand gesticulation, changes in pitch and a termination of a grammatical clause as turn-yielding signals.

Others have argued against the general model of a reactionary approach as, put simply, turn-transitions occur too quickly and turn-yielding signals occur too late within a speaker's utterance for the listener to simply react to an end-of-turn signal (Levinson and Torreira [2015], Riest et al. [2015]). (It is important to note that the `reactionary` approach is only dismissed in the context of the entire psycholinguistic model where both production and comprehension is considered.)

Sacks et al. [1974] pioneered the `predictive` approach and in their analysis of turn-taking argued that the observed speed of turn-transitions required some form of 'projection' with the production of language beginning prior to the end of a turn. This model of turn-taking is based off of separating speech into units, where one participant is the speaker, called `Turn Construction Units (TCU)` and immediately after completing a TCU a `Transition Relevance Place (TRP)` occurs that signals that a turn-transition (turn-shift) can occur. It is also important to note that a TRP does not always result in a turn-shift and a turn-shift does not always occur at a TRP. Nevertheles, every TRP is governed by a set of rules determing whether or not a TRP will result in a turn-shift:

1. The current speaker may select a new speaker during which the other participants act as listeners
2. If the current speaker does not select then any participant can self-select. The first to start gains the turn.
3. If no other party self-selects, the current speaker may continue.

The rules highlights an interesting property of turn-taking, first noted by Sacks et al. [1974] and verified by Ten Bosch et al. [2005], that intra-speaker gaps (gaps within the same utterance) are longer than inter-speaker gaps (gap resulting in a turn-shift) by around 25%.

Sacks et al. [1974] note that in order for a listener to project the end-of-turn than the speaker would have to construct their turns, with successive TCUs, in such a way that a turn transition is foreshadowed, showing that the turn is, in effect, winding down.

Some effort has been taken by Heldner and Edlund [2010] to critique the `predictive` approach. They argued that the systematic properties outlined within Sacks et al. [1974] are not consistent with observed data. This claim originates from the number of turn transitions above 200ms (41%-45% Heldner and Edlund [2010]) and for these turn transitions, listeners are reacting to silence or reacting to phrase-final prosodic information. Levinson and Torreira [2015] comprehesnibly dismisses the claim by citing it takes for a silence to be recongizable, the reaction to the silence and any production of speech, where they claim the process would take 550ms. In addition to this, Riest et al. [2015] point out that the presense of longer gaps could be explained by a speaker intetionally delaying a response when producing a 'dispreferred' response (Levinson [1983], Kendrick and Torreira [2015]). The existence of phrase-final cues that contribute to turn-taking however, is an area of debate that could explain the complexities within the organisation of turn-taking.

2.1.2 Turn-taking Cues

The question remains, what features of speech are relevant when predicting a TCU completion and, as such, when completing a turn? Prior research related to turn-yielding signals (Duncan [1972]), pointed out prosodic, syntactic and gestural features that coincide with turn-completion at an end-of-turn. Later work focussed on these turn-yielding signals and which signals contribute in a meaningful manner such that the listener is able to project a turn-completion. Most work has focussed on three aspects of conversation: syntactic, prosodic and pragmatic features. Gestural features (Duncan [1972]) and gaze (Kendon [1967]) have shown to be a useful part of turn-taking but findings in gaze have suggested these features are action dependent and more context-sensitive than other features (Clayman [2012]).

Although Sacks et al. [1974] left solving the question of how projection occurs they suggested that syntax to future research but they suggested that syntax and semantics contributed more due to the projectibility of syntactic units as compared to the projectibility of prosodic units.

The complex nature of turn-taking and the constraints imposed by language production means that turn-taking cues have to be early enough in order to determine a turn-completion and to generate some kind of response. With knowledge of the temporal requirements De Ruiter et al. [2006] tasked participants to predict a turn-completion both when intonational contours have been removed and then when lexicosyntactic information is removed. They found that performance was unaffected by the removal of intonational context but it was heavily affected by the removal of lexicosyntactic infomation concluding that lexicosyntactic information is necessary for turn-taking and tentaively sufficient for turn-taking while this is not the case for intonational information. This study was backed up by Magyari and De Ruiter [2012] that when particpants predicted the remaining part of a sentence they were more accurate in predicting a turn-completion which could be due to a listener predicting the content of the sentence, as such predicting the length of the remaining utterance. This belief was also highlighted by Pickering and Garrod [2013] who found that listeners actually imitate the speaker to determine intention and as such the content which is combined with the speaker's speaking rate to correctly time their own prepared utterance.

However, purely considering lexicosyntactic information may result in predicting turn-shifts too frequently due to the prevelance of syntactic completeness points (Ford and Thompson [1996]) within sentences. For example (taken from Ford and Thompson [1996]):

> V: And his knee was being worn/- okay/ wait./ It was bent/ that way/

Here '/' represents a syntactic completion points where an utterance is syntactically complete if "in its discourse context, it could be interpreted as a complete clause, that is, with an overt or directly recoverable predicate, without considering intonation or interactional import". With so many syntactic completion points it is important to determine the differences between those to find correct TRPs. Ford and Thompson [1996] did so by defining a point of syntactic, intonational and pramatic completness as a `Complex Transition Relevant Phrase (CRTP)` where 71% of CRTPs are turnshifts. In this instance, intonational completeness refers to a completion of an intonational unit containing a single coherent intonational contour (Du Bois and Schuetze-Coburn [1993]) and pragmatic completness refers to a point of intonational completness with a complete conversational action. The importance of pragmatics is seen in Pickering and Garrod [2013] due to its description and reliance on intention to predict a turn-completion.

However, it may be important to determine the differences in what may or may not be used in determining a turn-completion. In Pickering and Garrod [2013] speaker intention plays an important role in allowing the listener to ultimately predict the end of an utterance. The role of intention aligns well with what is can be described as ‘action‘ as in Levinson [2012] and ‘pragmatics‘ within Ford and Thompson [1996] (although their defintion considers intonation). As such the removal of lexicosyntactic information as in De Ruiter et al. [2006] and the resulting effects could be related to the overall removal of pragmatics.

Ford and Thompson [1996] theorised that TCUs and their partnering TRPs are a complex notion and as such multiple factors should be considered for predicting a turn completion. This belief was tested by Bögels and Torreira [2015] who also sought to refute the claim that intonation had no effect on turn-taking prediction by De Ruiter et al. [2006]. This was done by performing the same experiment but with instances of questions with equal syntactic completion points but different turn-shift locations (e.g. Are you a student? vs. Are you a student at university?). They found that in cases of syntactical ambiguitiy, lexicosyntactic information is not sufficient for turn-end projection and as such they claim intonation plays a role of disambiguation.

The findings of Bögels and Torreira [2015] also suggest that listeners look out for turn-taking cues that are present later on within an utterance. These turn-taking cues are both lexicosyntactic and intonational. This suggests a more complicated psycholinguistic model of turn-taking that is explored within Levinson and Torreira [2015] where the planning of content is begun once the intention of the content is determined. As such the content is prepared and with the perception of turn completion, whether by syntactic, pragmatic or prosodic completeness, the utterance is produced. Levinson and Torreira [2015] suggests that the listener may specifically look for syntactic and prosodic completness to determine when to speak.

2.2 Models for End-of-Turn Detection and Prediction

The tradition around conversational systems’ turn-taking ability is based on the existence of a silence threshold. In these models a turn is assumed to have been yielded by the current spekaer once some threshold has been past (around 650ms). However, as it is to be expected, this approach yields sluggish or possibly, misstaken interruptions. As discussed above, human-human turn-taking organisation is complicated and nuanced and as such the models generated should aim to be able to utilise the signals available in conversation.

2.2.1 Classification-based Models

Further research brought this idea into fruition with what could be interpreted as 'IPU-based' models. An `Interpausal Unit (IPU)` is a segmented part of continuous speech without silence exceeding a certain threshold (200ms). IPU-based models still undertake some form of silence detection, just with a shorter threshold than a pure-silence model, and after sufficient silence has been detected the model predicts whether the silence is a TRP or a non-TRP and as such whether the turn has been yielded by the speaker.

Naturally, these models resembled the natural progression of state-of-the art machine learning models moving from rule-based classifier Bell et al. [2001], to a decision-tree classifier Sato et al. [2002], Ferrer et al. [2002], Schlangen [2006], Meena et al. [2014], Raux and Eskenazi [2008] and then now onto deep learning architectures including the use of an LSTM RNN architecture Maier et al. [2017]. Each model uses a different set of features and found varying results on the effectiveness of various prosodic, lexicosyntactic and pragmatic features. Specifically Sato et al. [2002] and Meena et al. [2014] found that prosody did not contribute significantly to a decision while Ferrer et al. [2002] and Schlangen [2006] found that syntactic and prosodic features both contribute to turn-taking accuracy.

Models such as Sato et al. [2002], Schlangen [2006], Meena et al. [2014] specifically use silence thresholds that are fixed in size. As such, if the speaker yields their turn and if the model does not detect an end of turn then a state of silence may continue. As such other models such as Ferrer et al. [2002], Raux and Eskenazi [2008] incorporate silence length in order to continouly condition a response based on the time of silence and as such the longer after a pause the more likely that the turn has in fact been yielded. Raux and Eskenazi [2008], took this step further by also using turn-holding cues in order to condition the silence threshold so when more turn-holding cues are detected, the system will wait longer before considering a turn-shift event.

This process of monitoring speaker cues, to determine a turn-holding intention Raux and Eskenazi [2008], introduced the concept of a continuous model to monitor turn-taking. An approach which has been all the more feasible with advances in both deep learning architectures and more powerful feature extractors or pretrained features. Rather than taking on a traditional approach of classifying an utterance, the continuous model processes an utterance incrementally so that at any point the model is able to predict the liklihood of a turn-shift. The system bares more symmetry with our human-human interactions as the system could be able to project turn-completions, determine intent or action and generate an appropriate response.

2.2.2 Continuous Models

Another issue with previous approaches to turn-taking, namaly the classification approach, is the availability of data that is accurately annotated. As well as this, speech data can be noisy as noted by Sacks et al. [1974] overlapping speech is common but brief, and these sections of speech should not constitue a turn-shift and so this has to be annotated well in data. Recognising this issue, Skantze [2017] proposed a general, continuous turn-taking model, that was trained in a self-supervised manner. Self-supervised as the model is predicting the voice activity of separate speakers over the next two secondsand so it is able to predict a turn-shift based on this speech activity data. The model is also continuous in that it makes these predictions in 50ms intervals.

Others have also adopted the general LSTM approach Maier et al. [2017], Roddy et al. [2018a] to investigate the effectiveness of certain features. The general consensus is that both features in conjunction are required for superior performance however Roddy et al. [2018a] found that acoustic features are more beneficial and Maier et al. [2017] found that using only linguistic features was in fact worse than the baseline used. Roddy et al. [2018b] introduced a multiscale approach where linguistic and prosodic features are processed with different temporal speeds, linguistic features at slower time frames, which yielded better results than the previous baseline provided by both Roddy et al. [2018a], Skantze [2017] suggesting that the considering the two features at different granualities can provide additional benefits.

In order to seek improvements over these deep learning approaches it is imperative to discuss the quality of features in use. Prosodic features are generally a far more concrete cue to consider than linguistic features. Semantics of a conversation is difficult for models to discren as natural language is complex and ambiguous and any one intention can be represented in a variety of different word formations making overall semantics and pramatics generated by individual linguistic tokens difficult to capture. As such, linguistic feature representation within Roddy et al. [2018a], Skantze [2017], Maier et al. [2017] have been simplistic or in some cases non-existant Ward et al. [2018]. Skantze [2017] solely used POS tags. Roddy et al. [2018a] used a linear neural network to generate embeddings specific for turn-taking and Maier et al. [2017] used an enriched language model, trained by predicting a hidden word and in this case a special token representing the end of a turn.

This usage of a special token is a step undertaken by TurnGPT (Ekstedt and Skantze [2020]). Stonger language models allows for greater pragmatic and semantics and various contexts to be captured and by using these enhanced embeddings the predictive power of lingsuistic features will increase. TurnGPT is based on Open AI's GPT-2 (Radford et al. [2019]) which has been pretrained on a next-word prediction task and as such TurnGPT performs additional finetuning by the addition of speaker tokens and a special token indicicating the end of a turn. The model shows increased performance when considering context as well as a greater consideration of pragmatic completeness over its LSTM baseline.

Open AI's GPT-2 (Radford et al. [2019]) is based on the transformer decoder (Liu et al. [2018]) structure while `Bidirectional Encoder Representations from Transformers (BERT)` () is based primarily on the original transformer architecture (Vaswani et al. [2017]), the main difference is that BERT is able to utilize bidirectional self-attention wheras the GPT in a general sense can only be used in an autoregressive manner, so by attending to context on the left. BERT is trained based off of two tasks, Mask LM, where the model predicts 15% unknown tokens and next sentence predition.

Ekstedt and Skantze [2020] suggest that by generating responses using GPT-2's primary function of generating output, the system can predict how long it will be untill a turn-completion by generating multiplte possible outputs. This idea was furthered by Jiang et al. [2023] who proposed, RC-TurnGPT, which augments TurnGPT's training procedure by using the previous context and the next utterance in order to generate the probability of a turn shift. This extension of context allows for more considered turn-taking where the model does not take a turn at an early completion point. This could, for example, be a statement followed by a question. As such it explore the room of exploration that could be achieved by considering both directions of contextual history and future in producing a turn-shift prediction.

Chapter 3

Your next chapter

A dissertation usually contains several chapters.

Chapter 4

Conclusions

4.1 Final Reminder

The body of your dissertation, before the references and any appendices, *must* finish by page 40. The introduction, after preliminary material, should have started on page 1.

You may not change the dissertation format (e.g., reduce the font size, change the margins, or reduce the line spacing from the default single spacing). Be careful if you copy-paste packages into your document preamble from elsewhere. Some LaTeX packages, such as `fullpage` or `savetrees`, change the margins of your document. Do not include them!

Over-length or incorrectly-formatted dissertations will not be accepted and you would have to modify your dissertation and resubmit. You cannot assume we will check your submission before the final deadline and if it requires resubmission after the deadline to conform to the page and style requirements you will be subject to the usual late penalties based on your final submission time.

Bibliography

Linda Bell, Johan Boye, and Joakim Gustafson. Real-time handling of fragmented utterances. In *Proc. NAACL workshop on adaptation in dialogue systems*, pages 2–8, 2001.

Sara Bögels and Francisco Torreira. Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, 52:46–57, 2015.

Steven E Clayman. Turn-constructional units and the transition-relevance place. *The handbook of conversation analysis*, pages 151–166, 2012.

Jan-Peter De Ruiter, Holger Mitterer, and Nick J Enfield. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3):515–535, 2006.

John W Du Bois and Stephan Schuetze-Coburn. Representing hierarchy: Constituent structure for discourse databases. *Talking data: Transcription and coding in discourse research*, pages 221–260, 1993.

Starkey Duncan. Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, 23(2):283, 1972.

Starkey Duncan. Toward a grammar for dyadic conversation. 1973.

Starkey Duncan. On the structure of speaker–auditor interaction during speaking turns1. *Language in society*, 3(2):161–180, 1974.

Starkey Duncan and Donald W Fiske. *Face-to-face interaction: Research, methods, and theory*. Routledge, 2015.

Erik Ekstedt and Gabriel Skantze. Turngpt: a transformer-based language model for predicting turn-taking in spoken dialog. *arXiv preprint arXiv:2010.10874*, 2020.

Luciana Ferrer, Elizabeth Shriberg, and Andreas Stolcke. Is the speaker done yet? faster and more accurate end-of-utterance detection using prosody. In *Seventh international conference on spoken language processing*, 2002.

Cecilia E Ford and Sandra A Thompson. Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. *Studies in interactional sociolinguistics*, 13:134–184, 1996.

Mattias Heldner and Jens Edlund. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568, 2010.

Bing'er Jiang, Erik Ekstedt, and Gabriel Skantze. Response-conditioned turn-taking prediction. *arXiv preprint arXiv:2305.02036*, 2023.

Adam Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26:22–63, 1967.

Kobin H Kendrick and Francisco Torreira. The timing and construction of preference: A quantitative study. *Discourse Processes*, 52(4):255–289, 2015.

Stephen C Levinson. *Pragmatics*. Cambridge university press, 1983.

Stephen C Levinson. Action formation and ascription. *The handbook of conversation analysis*, pages 101–130, 2012.

Stephen C Levinson and Francisco Torreira. Timing in turn-taking and its implications for processing models of language. *Frontiers in psychology*, 6:731, 2015.

Peter J Liu, Mohammad Saleh, Etienne Pot, Ben Goodrich, Ryan Sepassi, Lukasz Kaiser, and Noam Shazeer. Generating wikipedia by summarizing long sequences. *arXiv preprint arXiv:1801.10198*, 2018.

Lilla Magyari and Jan P De Ruiter. Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in psychology*, 3:376, 2012.

Angelika Maier, Julian Hough, David Schlangen, et al. Towards deep end-of-turn prediction for situated spoken dialogue systems. 2017.

Raveesh Meena, Gabriel Skantze, and Joakim Gustafson. Data-driven models for timing feedback responses in a map task dialogue system. *Computer Speech & Language*, 28(4):903–922, 2014.

Martin J Pickering and Simon Garrod. An integrated theory of language production and comprehension. *Behavioral and brain sciences*, 36(4):329–347, 2013.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

Antoine Raux and Maxine Eskenazi. Optimizing endpointing thresholds using dialogue features in a spoken dialogue system. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 1–10, 2008.

Carina Riest, Annett B Jorschick, and Jan P de Ruiter. Anticipation in turn-taking: mechanisms and information sources. *Frontiers in psychology*, 6:89, 2015.

Matthew Roddy, Gabriel Skantze, and Naomi Harte. Investigating speech features for continuous turn-taking prediction using lstms. *arXiv preprint arXiv:1806.11461*, 2018a.

Matthew Roddy, Gabriel Skantze, and Naomi Harte. Multimodal continuous turn-taking prediction using multiscale rnns. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pages 186–190, 2018b.

H. Sacks, E.A. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. volume 50, page 696 – 735. 1974.

Ryo Sato, Ryuichiro Higashinaka, Masafumi Tamoto, Mikio Nakano, and Kiyoaki Aikawa. Learning decision trees to determine turn-taking by spoken dialogue systems. In *INTERSPEECH*, 2002.

David Schlangen. From reaction to prediction: Experiments with computational models of turn-taking. *Proceedings of Interspeech 2006, Panel on Prosody of Dialogue Acts and Turn-Taking*, 2006.

Gabriel Skantze. Towards a general, continuous model of turn-taking in spoken dialogue using lstm recurrent neural networks. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 220–230, 2017.

Louis Ten Bosch, Nelleke Oostdijk, and Lou Boves. On temporal aspects of turn taking in conversational dialogues. *Speech Communication*, 47(1-2):80–86, 2005.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Nigel G Ward, Diego Aguirre, Gerardo Cervantes, and Olac Fuentes. Turn-taking predictions across languages and genres using an lstm recurrent neural network. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 831–837. IEEE, 2018.

Victor H Yngve. On getting a word in edgewise. In *Papers from the sixth regional meeting Chicago Linguistic Society, April 16-18, 1970, Chicago Linguistic Society, Chicago*, pages 567–578, 1970.

Appendix A

First appendix

A.1 First section

Any appendices, including any required ethics information, should be included after the references.

Markers do not have to consider appendices. Make sure that your contributions are made clear in the main body of the dissertation (within the page limit).

Appendix B

Participants' information sheet

If you had human participants, include key information that they were given in an appendix, and point to it from the ethics declaration.

Appendix C

Participants' consent form

If you had human participants, include information about how consent was gathered in an appendix, and point to it from the ethics declaration. This information is often a copy of a consent form.