How to prepare the video component of the *Diachronic Corpus of Political Speeches* for multimodal analysis

Camille Debras
Université Paris Nanterre / France

Abstract – The *Diachronic Corpus of Political Speeches* (DCPS) is a collection of 1,500 full-length political speeches in English. It includes speeches delivered in countries where English is an official language (the US, Britain, Canada, Ireland) by English-speaking politicians in various settings from 1800 up to the present time. Enriched with semi-automatic morphosyntactic annotations and with discourse-pragmatic manual annotations, the DCPS is designed to achieve maximum representativeness and balance for political English speeches from major national English varieties in time, preserve detailed metadata, and enable corpus-based studies of syntactic, semantic and discourse-pragmatic variation and change on political corpora. For speeches given from 1950 onwards, video-recordings of the original delivery are often retrievable online. This opens up avenues of research in multimodal linguistics, in which studies on the integration of speech and gesture in the construction of meaning can include analyses of recurrent gestures and of multimodal constructions. This article discusses the issues at stake in preparing the video-recorded component of the DCPS for linguistic multimodal analysis, namely the exploitability of recordings, the segmentation and alignment of transcriptions, the annotation of gesture forms and functions in the software ELAN and the quantity of available gesture data.

Keywords – DCPS; multimodal political discourse analysis; gesture studies

1. INTRODUCTION

Still under construction, the *Diachronic Corpus of Political Speeches* (henceforth DCPS) is a collection of 1,500 full-length political speeches in English. It includes speeches delivered worldwide by English-speaking male and female elected politicians in various settings (such as election speeches, parliamentary or party conference speeches, inaugural addresses) from 1800 up to the present time. Speech transcripts are being enriched with semi-automatic morphosyntactic annotations, in the form of lemmatisation and part-of-speech tagging with *TreeTagger* (Schmid 1994). Transcripts are also supplemented with discourse-pragmatic manual annotations including audience responses (Heritage and Greatbatch 1986), speech openings and closures. The DCPS is

designed to fulfil the following criteria: achieve maximum representativeness and balance for political English speeches from major national English varieties in time, preserve detailed metadata, as well as enable corpus-based studies of syntactic, semantic and discourse-pragmatic variation and change on political corpora.

For speeches given from 1950 onwards, video-recordings of the original delivery are often retrievable online. This opens up avenues of research in multimodal linguistics, in which studies on the integration of speech and gesture in the construction of meaning (Kendon 2004; Norris 2004; Müller *et al*. 2013, *inter alia*) can include analyses of recurrent gestures (Müller *et al.* 2013) and of multimodal constructions (Steen and Turner 2013; Zima and Bergs 2017).

This article discusses the practical and methodological issues at stake in preparing the video-recorded component of the DCPS for linguistic multimodal analysis, including the analysis of gesture. It focuses on four main areas:

1) Exploitability of the recordings.
2) Preparing the spoken component of the data: transcription, segmentation of the transcription and alignment of the transcription with the video.
3) Preparing the visual component of the data: annotation of gesture forms and their functions.
4) Quantity of data: how much is "enough"?

After a short presentation of current issues in multimodal discourse analysis, I develop the issues at stake in these four areas in separate sections, before moving on to a general conclusion on the perspectives for innovative diachronic multimodal analyses of political discourse based offered by the DCPS.

2. MULTIMODAL (POLITICAL) DISCOURSE ANALYSIS

2.1. *A multimodal approach to the study of discourse: A focus on the interaction between speech and gesture*

Spoken communication is multimodal by nature (Norris 2004, *inter alia*). Meaning-making relies on the integration of multiple modes of communication which belong to two different modalities of communication: the oral-aural and the kinesic-visual modalities. The multimodal approach to (political) discourse analysis proposed here focuses on the interplay between actions in speech and gesture that are coordinated in

time (Kendon 2000; Mondada 2016). The contribution of speech can be further subdivided into the verbal mode (discourse, at the segmental level) and the vocal mode (at the suprasegmental level: prosodic phenomena including intonation, volume and speed of the delivery), alongside the gestural mode (Ferré 2011, 2019). Visible bodily actions, including gestures, are so closely intertwined with speech in the construction of meaning that they can be considered part of language themselves (Kendon 2000; Müller *et al.* 2013, *inter alia*). Alongside multimodality, a second main feature of spoken communication is sequentiality: actions done with speech or the body are inscribed in time, one after the other or simultaneously, and they take on their meanings and functions as part of this simultaneous and sequential unfolding of actions and mobilisation of resources. A gesture can thus be defined as a bodily action that "(belongs) to the 'story line' of the interaction" (Kendon 1986: 6), and that is inscribed in the sequentiality of the interaction, namely that coincides with other actions in the construction of meaning, rather than being there by mere coincidence (Schegloff 1984).

The study of gesture is fundamentally interdisciplinary (Stam and Ishino 2011). Since the analysis of a gesture lends itself to a large range of approaches in various domains (e.g. psychology or anthropology), a recent body of work has developed a specifically linguistic approach to gesture (cf. Müller *et al.* 2013), showing how gestures, traditionally relegated to the para-verbal, actually do lend themselves to linguistic analysis. Although not prototypically linguistic, gestures can be linguistic to some extent (Cienki 2017), notably in terms of forms, functions, and form-function pairings, for instance in the case of recurrent gestures (Ladewig 2014).

The multimodal study of political discourse is a thriving field, which includes the contribution of gestures to discourse structuring and framing (cf. Streeck 2008; Cienki 2009; Wehling 2009; Cienki and Giansante 2014; Debras and L'Hôte 2015, *inter alia*). The video component of the DCPS will provide scholars with new opportunities to supplement this field with studies on (diachronic) variation.

2.2. ELAN, a tool for the study of multimodal data

Several software tools can be used for the annotation and study of multimodal data. In this article, I choose to focus on ELAN[1] (Sloetjes and Wittenburg 2008), a professional,

---

[1] ELAN can be downloaded for free at https://tla.mpi.nl/tools/tla-tools/elan/. The current version available is version 6.1.

open-source tool for the creation of complex annotations on video resources developed at the Max Planck Institute for Psycholinguistics in Nijmegen (Netherlands), which is well-suited to the annotation and study of gesture. A detailed description of the software is available at the ELAN website,[2] which I reproduce below:

> With ELAN a user can add an unlimited number of annotations to audio and/or video streams. An annotation can be a sentence, word or gloss, a comment, translation or a description of any feature observed in the media. Annotations can be created on multiple layers, called 'tiers'. Tiers can be hierarchically interconnected. An annotation can either be time-aligned to the media or it can refer to other existing annotations. The textual content of annotations is always in Unicode and the transcription is stored in an XML format.

ELAN provides different views on the annotations, and each view is connected and synchronised to the media play head. Up to four video files can be associated with an annotation document. Each video can be integrated in the main document window or displayed in its own resizable window. ELAN delegates media playback to an existing media framework, like *Windows Media Player*, *QuickTime* or *Java Media Framework* (JMF). As a result, a wide variety of audio and video formats is supported, and high-performance media playback can be achieved. ELAN is written in the *Java* programming language and the sources are available under a GPL 3 license. It runs on Windows, Mac OS X and Linux. ELAN's main other features are:

- Navigate through the media with different step sizes.
- Easy navigation through existing annotations.
- Waveform visualisation of .wav files.
- Support for template documents.
- Input methods for a variety of script systems.
- Multi-tier regular expression search, within a single document or in a selection of annotation documents.
- Support for user definable Controlled Vocabularies.
- Import and export of Shoebox/Toolbox, CHAT, Transcriber (import only), *Praat* and .csv/tab-delimited text files.
- Export to interlinear text, html, smil and subtitles text.
- Printing of the annotations.
- Multiple undo/redo.

---

[2] https://tla.mpi.nl/tools/tla-tools/elan/elan-description/

ELAN is an especially convenient annotation tool because it is open source, compatible with other video and audio transcription and annotation software like CLAN[3] (MacWhinney 2000) or *Praat*[4] (Boersma and Weeninck 2017), and because the annotated data can easily be exported for further analysis or statistical calculus in .csv or tab-delimited format.

2.3. What counts as "the data"

One epistemological issue raised by the study of video-recordings of political speeches is the status of the recording with respect to what counts as the data. If the data is primarily considered to be the speech as it was delivered in its original setting, in the co-presence of its addressees, then the videorecording can be seen as a mere tool for accessing the data itself. This view is, for instance, usually adopted in the field of Conversation Analysis to analyse talk-in-interaction (Mondada 2009). And yet, with the advent of media culture (television, the Internet), some recordings are arguably also designed first and foremost as mediated broadcasts. In this sense, the videorecording itself can also be regarded as primarily constitutive of the data. In a diachronic corpus like the DCPS, this second view will probably be increasingly relevant with time and with the development of mass media communication. Both views can, of course, hold true at once, with the speech being designed both for an in-person and online audience, both as an interaction and as a mediated broadcast, and, indeed, the speech as a successful interaction can contribute to its success as a broadcast.

3. EXPLOITABILITY OF THE VIDEO-RECORDING FOR MULTIMODAL ANALYSIS

For speeches given from 1950 onwards that will be included in the DCPS, video-recordings of the original delivery are retrievable online —often for free on video-sharing platforms like *YouTube*. Original videos can usually be downloaded with free online tools that convert *YouTube* videos into .mp4 video clips that can be stored on a computer or hard drive.[5] To my knowledge, there are no ethics or copyright issues

---

[3] http://dali.talkbank.org/clan/

[4] http://www.fon.hum.uva.nl/praat/

[5] For instance https://youtube-mp4.download/fr/free-converter, or the *Firefox Extension Downloadhelper*, https://addons.mozilla.org/fr/firefox/addon/video-downloadhelper/

related to collecting and analysing data for research purposes that are already freely available on online platforms.

Latest versions of ELAN are compatible with several video formats. Yet .mp4 is a sound choice, as a common widespread format, it can be played easily on either a Mac or a PC and, as a compressed format, it will allow for slightly lighter video files with enough quality of recording. A minimum quality of recording will be needed for hand gestures, posture changes, head movements and facial expressions to be visible to the human eye. Compressed formats can, nevertheless, impact the quality of the recordings and, in turn, hinder visual analysis. If multiple recordings of the same speech are available online, priority can be given to the less compressed format, if possible, as well as to the characteristics that will be discussed in what follows (cf. Sections 3.1–3.3.)

3.1. Camera framing

Political speeches are often filmed with a fairly close framing that leaves most of the hand gestures out (medium close-up framing; cf. Figure 1). If a recording with a larger framing is available (medium shot, from the waist up; cf. Figure 2), it can be preferred so as to capture as many gestures as possible. And yet, if only close-up framings are available, they are of course relevant for multimodal analysis as well. Indeed, political orators are often used to medium close-up framings and adapt by relying mostly on gestures that are visible on camera, such as facial expressions, head and shoulder movements, and hand gestures with a small amplitude or realised in the visible part of the gesture space.
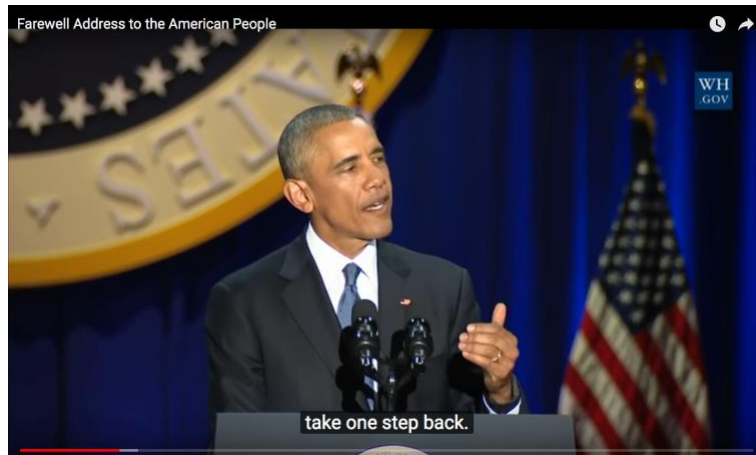

Figure 1: Medium close-up framing

Figure 2: Medium shot framing

3.2. Camera angle

A recording that faces the speaker should be preferred to one with a camera positioned sideways. And yet, if the speaker is positioned behind a high reading desk, gestures might be partly hidden behind it. In that case, if a recording with a sideways camera angle is available, it can be preferred so as to capture more of the speaker's hand gestures (cf. Figure 3).

Figure 3: Sideways angle with high desk

3.3. Continuity of filming

Discontinuous camera shooting is frequent. Editing choices can include large camera framings of the audience or zoom-ins on certain audience members. Such interventions can result in gestures of the speaker being realised out of frame. If multiple recordings of the same speech are available, the most continuous recording should be preferred. If the most continuous (or the only available) recording potentially includes out-of-frame

gestures (e.g. framings of the audience while the speech is heard as a voice-over), the recording remains exploitable for multimodal analysis. In such (frequent) cases, the work will bear on whatever occurrences are observable.

4. TRANSCRIPTION, SEGMENTATION AND ALIGNMENT

To navigate multimodal data, the transcription of each speech must be segmented into units, and these units need to be temporally aligned with the stream of the corresponding video. When transcripts are available online, for instance on institutional websites, they are usually fairly accurate. It can happen that some aspects typical of spoken delivery have been smoothed out in written transcripts. If so, the transcript should be proofread and corrected so as to reintroduce the marks of orality (e.g. hesitations, filled pauses, discourse markers, repetitions, repairs) that are actually produced by the speaker. This would be useful so as to carry diachronic studies of the emergence of the conversational framing in political discourse (Cienki and Giansante 2014). Various units can be envisaged for the segmentation of speech. I choose to focus on two of them: the interpausal unit (automatic segmentation) and the intonation unit (manual segmentation).

4.1. The interpausal unit (IPU)

IPUs are defined, in the framework of Conversation Analysis, as blocks of speech separated by silent pauses of 0.2 seconds (Koiso *et al*. 1998). The length of the pause can vary according to the language used and to the speech situation. Since monological speech during public political address is fundamentally different from spontaneous conversation (Rossette-Crake 2019, *inter alia*), the length of separation pauses will probably need to be readjusted for some speakers in the DCPS. The automatic alignment and segmentation into IPUs of the transcript can be realised automatically, for instance, with the software *SPeech Phonetization Alignment and Syllabification* (SPPAS; Bigi 2012; Bigi and Hirst 2012).[6] SPPAS can also produce automatic annotations of word, syllable and phoneme segmentations from a recorded speech sound and its corresponding transcription. The resulting alignments are a set of *TextGrid* files, the native file format of the *Praat* software, which will need to be

---

[6] http://www.sppas.org/

corrected manually. They can then be converted into annotation tiers in the ELAN software, where additional gesture annotations will be made.

4.2. The intonation unit (IU)

If there is opportunity (e.g. time, funding) for finer corpus annotation work, the DCPS will be manually segmented into IUs (Chafe 1994), also known as 'intermediate phrases' (Pierrehumbert and Hirschberg 1990), 'intonation phrases' (Wells 2006), or 'tone-units' (Lelandais and Ferré 2019), which can be considered as the spoken equivalent of the clause, as theorised in the well-established tradition of the British school of intonation analysis (Halliday 1967; Wells 2006, *inter alia*). An intonation phrase is organised around at least one nucleus, and characterised by a dynamic pitch contour, and is considered to constitute an information unit (Chafe 1994).

Chafe (1994: 58) defines six characteristics of prototypical IUs, namely pitch, duration, intensity, pausing, voice quality and speaker turn. Pitch (that is, fundamental frequency) usually includes a resetting of the pitch baseline (as in a 'step- up' or 'step-down' in the pitch level) and a recognisable final pitch contour (e.g., falling or rising). Duration usually includes increased tempo at the beginning (as in a shortening of syllables and/or words), and then a gradual slowing down toward the end (as in a lengthening of syllables and/or words). Intensity usually includes one or more syllables and/or words spoken more loudly. Pausing is often preceded or followed by pausing (but may also contain pauses within its boundaries). Voice quality sometimes begins or ends with a creaky voice or whispering. Finally, speaker turn may sometimes be associated with a change of speaker.

Accordingly, the prototypical intonation phrase can be defined as follows:

> (…) it is a spate of talk delivered as one recognisable overall pitch movement. In a standard textbook scenario this pitch movement would contain a pitch accent near the beginning, and another, typically more prominent pitch accent on the final stressed syllable; it would start with a comparatively high pitch onset, which would be followed by gradual declination in overall pitch register and loudness; the last syllable would be lengthened; and the whole phrase would be followed by a brief pause (Szczepek Reed 2011: 351).

Since the annotation of intonation units needs to be done manually, notably based on acoustic analyses of pitch contours in *Praat*, inter-coder reliability will need to be

established through a statistical test, such as Cohen's Kappa (McHugh 2012), so as to ensure the robustness of the annotation. As with any form of manual annotation, differences between annotators may occur: less experienced coders can, for instance, miss some boundaries. Ideally, a section of the data will be transcribed by at least two coders: experienced and less experienced ones. That way, less experienced coders may improve their annotation skills from confronting several transcripts. I here refer the reader to Stelma and Cameron's (2007) methodological paper on building skills for Intonation Unit annotation.

5. GESTURE ANNOTATIONS

5.1. Proposed guidelines for the formal and functional annotation of gesture

Although encouraging progress is being made in the automatisation of gesture annotation, notably thanks to motion capture technologies, most gesture annotation still has to be done manually. Gesture annotation is a time-consuming process. As explained in Section 3, a medium close-up framing is often favoured when politicians are filmed during monological public address. Therefore, the most visible gestures, which are usually also the ones that are the most mobilised by these coached public speakers, are facial expressions, head and shoulder movements, as well as hand gestures with a small amplitude or realised in the visible part of the gesture space.

Several very thorough reliable annotation systems have been developed for the annotation of gestures, such as the *Facial Action Coding System* (FACS; Ekman and Rosenberg 1997), for the annotation of facial expressions, and the *Linguistic Annotation System for Gestures* (LASG; Bressem *et al.* 2013) for the annotation of hand gestures. The main downside with these systems is precisely the direct consequence of their robustness and quality: they demand a lot of annotation time because they are extremely detailed, and since they require a significant degree of expertise and practice, they are not very adapted to beginners or non-specialists. Indeed, as Waller and Pasqualini (2013: 920) explain, the mastery of FACS is quite demanding:

> To use the system, researchers must learn to identify these base units of facial movements using a detailed manual (Ekman et al. 2002) and take a final test for certification. Although "in-house" inter-coder reliability may be desirable for specific studies in addition to

certification, it is not recommended as a substitute. Training takes an estimated 100 hours but can take more or less time depending on the context.

Likewise, LASG, which proposes the intonation unit as the unit of analysis for speech (cf. Section 4.2), constitutes a comprehensive yet quite complex analysis of the linguistic co-expressiveness of speech and gesture (McNeill 1992), as shown in Table 1.

Level of annotation		Name of Tier	obligatory/ optional	controlled vocabulary
Annotation of gestures	determining units	Gesture Unit	obligatory	
		Gesture Phases		
	annotation of form	Hand Shape	obligatory	
		Orientation		
		Position		
		Movement Type		
		Movement Direction		x
		Movement Quality		
	motivation of form	Mode of representation (MoR)	obligatory	
		Action		
		Motor pattern		
		Image schema		
Annotation of speech	annotation of speech (turn)	Speech Turn	obligatory	
		Speech Turn-translation		
		Speech Turn-Gesture Phases		
		Speech Turn-Gesture Phases translation		
	annotation of speech (intonation unit)	Intonation Unit		
		Intonation Unit-translation		
		Intonation Unit-Gesture Phases		
		Intonation Unit-Gesture Phases translation		
Annotation of gestures in relation to speech	prosody	Final pitch movement	obligatory	x
		Accent (primary, secondary)	optional	
	Syntax	Word Class		x
		Syntactic Function	optional	
		Integration		
	Semantics	Temporal Relation	obligatory	
		Semantic Relation	optional	x
		Semantic Function		
	Pragmatics	Turn	obligatory	
		Speech Act		
		Pragmatic Function	optional	x
		Dynamic Pattern		

Table 1: Overview of the levels of annotation in the LASG system (from Bressem *et al.*, 2013: 1101)

If preparatory gesture annotations can be made on the DCPS in ELAN, in the perspective of multimodal analysis, they do not need to be as fine-grained as these systems require. Depending on the annotation resources available, I suggest a series of more coarse-grained preparatory annotations that are accessible to less expert coders. These coding recommendations are rooted in a form-based approach (Boutet 2008; Müller *et al.* 2014), and partly inspired from the MUMIN annotation scheme (Allwood *et al.* 2005) and from the LASG system. Indeed, as Bressem *et al.* (2013: 1104) note:

> The Linguistic Annotation System for Gestures approaches the description of gestures' forms by applying the four parameters "hand shape", "orientation", "movement" and "position in gesture space", developed for the description of signs (Battison 1974; Stokoe 1960) to gestures. Taking the four form parameters as the basis of a gestural form description aims at systematically addressing the form aspects of a gestural Gestalt. In doing so, it allows for a fine-grained description of gestures and for a detection of gestural patterns and structures (e.g. […] Kendon 2004; Müller 2004).

If all the preparatory annotations suggested below cannot be made due to time or material constraints, we propose an order of priority to realise them. First, annotators will make formal annotations indicating what articulator is used (e.g. head, hand, eyebrows) and what configuration of the articulator (gesture form) is used, (e.g. for the head: head nod, head shake, head tilt). Second, functional annotations will be made, indicating the co-verbal function of hand gestures (McNeill 1992; Kendon 2004) and head movements (McClave 2000; Kendon 2002).

An annotation line or tier will be created in ELAN for each articulator used by the speakers. The annotation of a gesture form will include all gestures phases (cf. Kendon 2004): extension (i.e. departure from the resting or 'home' position (Sacks and Schegloff 2002), the stroke, the hold if applicable, and the retraction phase until the return to the home position, or the shift to another gesture form. For instance, in the case of a palm-up open-hand (PUOH) gesture, the annotation will start from the extension of the hand from the resting position to the gesture stroke itself during which the palm is turned upward (stroke) and possibly held in that position (gesture hold), until the hand reaches back the home position after total retraction. Gesture phases do not need to be annotated right away: their detailed annotation can be done in a dependent tier later on by a gesture researcher, if the multimodal focus of the study requires it. Each annotation can be labelled with a description of the gesture form (configuration of the body

articulator involved), according to the annotation tables proposed below. Table 2 proposes a formal annotation of movements of the shoulders (which are often connected with shrugging (see Kendon 2004; Streeck 2008; Debras 2017), and of the face, inspired from the MUMIN annotation guide (Allwood *et al*. 2005). In turn, Tables 3 and 4 propose labels for the formal and functional annotation of head movements and hand gestures respectively. Tables 2 and 3 both include references about the chosen labels in the rightmost column, should the reader wish to seek more information about them.

Annotation tier	Annotation labels
Shoulders	Both shoulders lifted
	Right shoulder lifted
	Left shoulder lifted
	Other
Mouth/	Smile
Lips	Laughter
	Corners up
	Corners down
	Protruded
	Retracted
	Other
Gaze	Up
	Down
	Sideways
	Other
Eyes	Exaggerated Opening
	Closing-both
	Closing-one
	Closing-repeated
	Other
Eyebrows	Frowning
	Raising
	Other

Table 2: Labels for the annotation of non-manual gestures (with adaptations from Allwood *et al*. 2005)

Form (after Allwood *et al*. 2005)	Function	Corresponding reference for further information on the chosen labels
Single Nod (Down)	Assessment	Goodwin and
Repeated Nods (Down)	Agreement	Goodwin (1992)
Single Jerk (Backwards Up)	Inclusivity	
Repeated Jerks (Backwards Up)	Intensification	
Single Slow Backwards Up	Uncertainty	
Move Forward	Direct quotes	
Move Backward	Expression of mental images of characters	McClave (2000)
Single Tilt (Sideways)	Deixis and referential us of space	
Repeated Tilts (Sideways)	Lists or alternatives	
Side-turn	Lexical repair	
Shake (repeated)	Backchannelling request	
Waggle	Negation	Kendon (2002
Other	Other	

Table 3: Labels for the formal and functional annotation of head movements

	Annotation tier	Annotation labels	Corresponding reference for further information on the chosen labels
Main annotation tier for hand gestures	Hand	Both hands Left hand Right hand	
Formal and functional annotation tiers that are hierarchically dependent on the main annotation tier	Handshape (See illustrations below)	Index pointing Precision grip Vertical Palm PUOH (Palm-Up Open Hand) Thumb pointing Other	Kendon (2004) Müller (2004)
	Orientation	Vertical palm facing outwards Vertical palm facing sideways Horizontal palm down Horizontal palm sideways Horizontal palm up Oblique Other	
	Trajectory	Up Down Sideways Complex Other	Allwood *et al.* (2005)
	Localisation in the gesture space	Centre Periphery	
	Semantic relation with speech	Redundant Complementary/Supplementary Contrary Replacing	Bressem *et al.* (2013: 1111)
	Mode of representation (for iconic gestures only)	Drawing Molding Representing Acting	Müller (2014)
	Main function with respect to speech	Iconic Deictic Metaphoric Beat Emblem Interactive Recurrent gesture	McNeill (1992) Bavelas *et al.* (1992) Ladewig (2014)

Table 4: Labels for the formal and functional annotation of hand gestures

5.2. *Some recurrent gesture forms in political oratory*

Figures 4 to 7 present illustrations of handshapes that are recurrently used by political orators. A remarkable handshape that is used only in political oratory, by orators like Bill Clinton, Barack Obama or Justin Trudeau, is thumb pointing, also known as the 'Clinton thumb' (Mankiewicz 2006), although John F. Kennedy was first observed using it, presented in Figure 4. When they want to point to an abstract idea (abstract deixis, as per McNeill *et al.* 1993) and/or stress a point with a prosodic beat gesture (McNeill 1992), certain politicians are coached to point outward with the thumb rather

than with the index, so as to avoid the more aggressive connotations attached to finger pointing. Figure 5 shows an example of another handshape, index pointing, used with two distinct orientations.

Figure 4: Illustrations of thumb pointing

Figure 5: Index pointing with different orientations: horizontal palm down or upward

Figure 6 illustrates the 'precision grip' (Kendon 2004), a recurrent gesture whose core meaning is the expression of a precise, specific idea. Figure 7 is a typical example of a metaphoric gesture: the speaker is talking about an abstract referent, the topic of 'values', while representing this referent as a concrete object that can be manipulated — see also Streeck (2008) on speech-handling gestures.

Figure 6: Precision grip

Figure 7: Metaphoric gesture

5.3. Quantity of gesture

The multimodal component of the DCPS will provide at least several hundred kinesic forms. This is largely sufficient to conduct systematic linguistic analyses of gestures, notably of recurrent gestures and/or multimodal constructions (Steen and Turner 2013), notably with multivariate exploratory statistics tools in *R* (Desagulier 2017).

6. CONCLUSION

The DCPS is a collection of 1,500 full-length political speeches in English. This article aimed to present the practical and methodological issues at stake in preparing the video-recorded component of the DCPS for linguistic multimodal analysis, including the analysis of gesture, and to propose relevant recommendations and guidelines. It focused on four main issues: 1) the exploitability of recordings, 2) the segmentation and alignment of transcriptions, 3) the annotation of gesture forms and functions in ELAN and 4) the quantity of available gesture data. My main recommendations have the following:

1) Favour recordings with a medium shot framing so as to increase the visibility of the hand gestures used alongside facial displays and head movements.

2) Transcribe and segment the corpora in intonation units and rely on inter-coder agreement when segmenting the data.

3) Annotate gesture forms and functions based on the guidelines available in Tables 2, 3 and 4, which although already quite detailed are still coarser grained than existing annotation systems like FACS or LASG, and more accessible to less expert coders.

REFERENCES

Allwood, Jens, Loredana Cerrato, Kristiina Jokinen, Constanza Navarretta and Patrizia Paggio. 2005. The MUMIN annotation scheme for feedback, turn management and sequencing. *Proceedings of the 2nd Nordic Conference on Multimodal Communication*. Gothenburg, Sweden. http://www.sskkii.gu.se/jems/publications/bfiles/B80-3.pdf. (28 June, 2021.)

Battison, Robin 1974. Phonological deletion in American sign language. *Sign Language Studies* 5: 1–19.

Bavelas, Janet Bavelas, Nicole Chovil, Douglas A. Lawrie and Allan Wade. 1992. Interactive gestures. *Discourse Processes* 15/4: 469–489.

Bigi, Brigitte. 2012. SPPAS: A tool for the phonetic segmentation of speech. In Calzolari, Nicoletta, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asunción Moreno, Jan Odijk and Stelios Piperidis eds. *Proceedings of the 8th International Conference on Language Resources and Evaluation*, LREC 2012. Istanbul: European Language Resources Association, 1748–1755.

Bigi, Brigitte and Daniel Hirst. 2012. *SPeech Phonetization Alignement and Syllabification* (SPPAS): A tool for the automatic analysis of speech prosody. In Qiuwu Ma, Hongwei Ding and Daniel Hirst eds. *Proceedings of 6th Speech Prosody International Conference*. Shanghai: Tongji University Press, 19–22.

Boersma, Paul and David Weenink. 2017. *Praat: Doing Phonetics by Computer.* Computer Program. Version 6.0.28.

Boutet, Dominique. 2008. Une morphologie de la gestualité: Structuration articulaire. *Cahiers de Linguistique Analogique* 5: 81–115.

Bressem, Jana, Silva H. Ladewig and Cornelia Müller. 2013. Linguistic Annotation System for Gestures (LASG). In Cornelia Müller *et al*. eds., 1098–1125.

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

Cienki, Alan. 2009. Spoken language framing in political discourse. Presentation at the European Consortium for Political Research (ECPR). *Workshop Studying the Political through Frame Analysis,* 14–19 April 2009. Lisbon: Portugal.

Cienki, Alan. 2017. Language as a prototype category. In Alan Cienki ed. *Ten Lectures on Spoken Language and Gesture from the Perspective of Cognitive Linguistics: Issues of Dynamicity and Multimodality*. Leiden: Brill, 163–182.

Cienki, Alan and Gianluca Giansante. 2014. Conversational framing in televised political discourse: A comparison from the 2008 elections in the United States and Italy. *Journal of Language and Politics* 13/2: 255–288.

Debras, Camille. 2017. The shrug: Forms and functions of a compound enactment. *Gesture* 16/1: 1–34.

Debras, Camille and Émilie L'Hôte. 2015. Framing, metaphor and dialogue – A multimodal approach to party conference speeches. *Metaphor and the Social World* 5/2: 177–204.

Desagulier, Guillaume. 2017. *Corpus Linguistics and Statistics with R. Introduction to Quantitative Methods in Linguistics*. New York: Springer.

Ekman, Paul and Erika Rosenberg eds. 1997. *What the Face Reveals. Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford: Oxford University Press.

Ekman, Paul, Wallace V. Friesen and Joseph C. Hager. 2002. *Facial Action Coding System*. Salt Lake City: Research Nexus.

Ferré, Gaëlle. 2011. Functions of three open-palm hand gestures. *Multimodal Communication* 1/1: 5–20.

Ferré, Gaëlle. 2019. *Analyse de Discours Multimodale. Gestualité et Prosodie en Discours*. Grenoble: Éditions de l'Université Grenoble Alpes.

Goodwin, Charles and Marjorie Harness Goodwin. 1992. Assessments and the construction of context. In Alessandro Duranti and Charles Goodwin eds. *Rethinking Context*. Cambridge: Cambridge University Press, 147–190.

Heritage, John and David Greatbatch. 1986. Generating applause: A study of rhetoric and response at party political conferences. *American Journal of Sociology* 92/1: 110–157.

Halliday, Michael Alexander Kirkwood. 1967. *Intonation and Grammar in British English*. The Hague: Mouton de Gruyter.

Kendon, Adam. 1986. Some reasons for studying gesture. *Semiotica* 62/1–2: 3–28.

Kendon, Adam. 2000. Language and gesture: Unity or duality? In David McNeill ed., *Language and Gesture*. Cambridge: Cambridge University Press, 47–63.

Kendon, Adam. 2002. Some uses of the head shake. *Gesture* 2/2: 147–182.

Kendon, Adam. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Koiso, Hanae, Yasuo Horiuchi, Akira Ichikawa and Yasuharu Den. 1998. An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Language and Speech* 41/3–4: 295–321.

Ladewig, Silva H. 2014. Recurrent gestures. In Cornelia Müller *et al*. eds., 1558–1574.

Lelandais, Manon and Gaëlle Ferré. 2019. The verbal, vocal and gestural expression of (in)dependency in two types of subordinate constructions. *Journal of Corpora and Discourse Studies* 2: 117–143.

MacWhinney, Brian. 2000. *The CHILDES Project: Tools for Analyzing Talk.* Mahwah: Lawrence Erlbaum Associates.

Mankiewicz, Josh. 2006. *For Politicians, the Gesture's the Thing*. https://www.nbcnews.com/id/wbna15609023. (28 June, 2021.)

McClave, Evelyn. 2000. Linguistic functions of head movements in the context of speech. *Journal of Pragmatics* 32/7: 855–878.

McHugh, Mary L. 2012. Interrater reliability: The kappa statistic. *Biochemia Medica* 22/3: 276–282.

McNeill, David. 1992. *Hand and Mind. What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

McNeill, David, Justine Cassell and Elena Levy. 1993. Abstract deixis. *Semiotica* 1/2: 5–19.

Mondada, Lorenza. 2009. Video recording practices and the reflexive constitution of the interactional order: some systematic uses of the split-screen technique. *Human Studies* 32/1: 67–99.

Mondada, Lorenza. 2016. Multimodal resources and the organization of social interaction. In Andrea Rocci and Louis de Saussure eds. *Verbal Communication*. Berlin: De Gruyter, 329–350.

Müller, Cornelia. 2004. Forms and uses of the palm up open hand: A case of a gesture family? In Cornelia Müller and Roland Posner eds., *The Semantics and Pragmatics of Everyday Gestures*. Berlin: Weidler Buchverlag, 233–256.

Müller, Cornelia. 2014. Gestural modes of representation as techniques of depiction. In Cornelia Müller *et al*. eds., 1687–1701.

Müller, Cornelia, Jana Bressem and Silva H. Ladewig. 2014. Towards a grammar of gestures: A form-based view. In Cornelia Müller *et al*. eds., 707–732.

Müller, Cornelia, Silva H. Ladewig and Jana Bressem. 2013. Gesture and speech from a linguistic perspective: A new field and its history. In Cornelia Müller *et al.* eds., 55–81.

Müller, Cornelia, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill and Sedinha Teßendorf eds. 2013. *Body – Language – Communication Vol. 1*. Berlin: Walter de Gruyter.

Müller, Cornelia, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill and Sedinha Teßendorf eds. 2014. *Body – Language – Communication Vol. 2*. Berlin: Walter de Gruyter.

Norris, Sigrid. 2004. *Analyzing Multimodal Interaction: A Methodological Framework*. New York: Routledge.

Pierrehumbert, Janet and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan and Martha E. Pollack eds. *Intentions in Communication*. Cambridge: MIT Press, 271–311.

Rossette-Crake, Fiona. 2019. *Public Speaking and the New Oratory: A Guide for Non-native Speakers*. London: Palgrave Macmillan.

Sacks, Harvey and Emanuel A. Schegloff. 2002. Home position. *Gesture* 2/2: 133–146.

Schegloff, Emanuel A. 1984. On some gestures' relation to talk. In J. Maxwell Atkinson and John Heritage eds. *Structures of Social Action*. Cambridge: Cambridge University Press, 266–296.

Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. *Proceedings of the International Conference on New Methods in Language Processing*. Manchester, United Kingdom. https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf. (28 June, 2021.)

Sloetjes, Hans and Peter Wittenburg. 2008. Annotation by category – ELAN and ISO DCR. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk and Daniel Tapias eds. *Proceedings of the 6th International Conference on Language Resources and Evaluation*, LREC 2008. Marrakech: European Language Resources Association, 816–820.

Stam, Gale and Mika Ishino eds. 2011. *Integrating Gestures: The Interdisciplinary Nature of Gesture*. Amsterdam: John Benjamins.

Steen, Francis and Mark Turner. 2013. Multimodal Construction Grammar. In Mike Borkent, Barbara Dancygier and Jennifer Hinnell eds. *Language and the Creative Mind*. Chicago: University of Chicago Press, 255–274.

Stelma, Juurd H. and Lynne J. Cameron. 2007. Intonation units in spoken interaction: Developing transcription skills. *Text and Talk* 27/3: 361–393.

Stokoe, William. 1960. *Sign Language Structure*. Buffalo: Buffalo University Press.

Streeck, Jürgen. 2008. Gesture in political communication: A case study of the democratic presidential candidates during the 2004 primary campaign. *Research on Language and Social Interaction* 41/2: 154–186.

Szczepek Reed, Beatrice. 2011. Units of interaction: 'Intonation phrases' or 'turn constructional phrases'? In Hi-Yon Yoo and Élisabeth Delais-Roussarie eds. *Proceedings of the International Conference on Prosody-Discourse Interface*, 351–363.

Waller, Bridget and Marcia Pasqualini. 2013. Analysing facial expression using Facial Action Coding Systems (FACS). In Cornelia Müller *et al*. eds., 917–931.

Wehling, Elisabeth. 2009. Argument is gesture war. Function, form and prosody of discourse structuring gestures in political argument, *Proceedings of the 35th Annual Meeting of the Berkeley Linguistics Society*: 54–65.

Wells, John C. 2006. *English Intonation: An Introduction*. Cambridge: Cambridge University Press.

Zima, Elisabeth and Alexander Bergs eds. 2017. Special issue: Towards a multimodal construction grammar. *Linguistic Vanguard* 3/1.

Corresponding author
Camille Debras
Centre de Recherches Anglophones
200 Avenue de la République
Université Paris Nanterre
92001 Nanterre Cedex
France
Email: cdebras@parisnanterre.fr