Encoding Interruptions in Parliamentary Data: From Applause to Interjections and Laughter

Tanja Wissik

Electronic version
URL: https://journals.openedition.org/jtei/4214
DOI: 10.4000/jtei.4214
ISSN: 2162-5603

Publisher
TEI Consortium

Electronic reference

For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.
Encoding Interruptions in Parliamentary Data: From Applause to Interjections and Laughter

Tanja Wissik

ABSTRACT

Parliamentary data, especially parliamentary discourse, is of interest to researchers from various fields in the humanities and social sciences. The growing number of machine-readable and annotated parliamentary text corpora opens up this field for computer-based quantitative analysis. In this paper I aim to give an overview of how parliamentary interruptions are recorded in official parliamentary records and how they are modeled and encoded in currently available machine-readable parliamentary corpora. Furthermore, I will discuss whether these encodings are suitable for computer-based quantitative analysis of parliamentary interruptions. I will suggest detailed encodings of parliamentary interruptions in TEI as an extension to the Parla-CLARIN
recommendations, to enable the extraction of parliamentary interruptions and to facilitate computerized quantitative analysis based on these encodings. As an example, I will use the encoding of interruptions in the Austrian Parliamentary Records.

INDEX

**Keywords:** parliamentary data, parliamentary records, parliamentary interruptions, encodings of parliamentary interruptions

ACKNOWLEDGEMENTS

This work was supported in part by a CLARIN Mobility Grant and the DYLEN Project funded by the ÖAW goldigital Next Generation Grant (GDNG 2018–020).

1. **Introduction**

Parliamentary data, especially parliamentary discourse, is of interest to researchers from various fields in the humanities and social sciences, such as linguistics, political sciences, communication studies, history, and law studies. In most countries, official parliamentary records are available in electronic format on the parliamentary websites and there is a growing number of machine-readable and annotated parliamentary corpora (see Fišer, Lenardič, and Erjavec 2018, 1321) that are opening up parliamentary discourse for computer-based quantitative analysis. Interesting research into parliamentary data includes studies related to parliamentary interruptions. However, most of the studies produced so far are of a qualitative nature.

In this paper I will give an overview of how parliamentary interruptions are recorded in official parliamentary records and how they are modeled and encoded in currently available machine-readable parliamentary corpora. Furthermore, I will discuss whether these encodings are suitable for computer-based quantitative analysis of parliamentary interruptions. I will suggest detailed encodings of parliamentary interruptions as an extension to the Parla-CLARIN recommendations (Erjavec and Pančur 2022) to facilitate computerized quantitative analysis based on these encodings. Such encodings would allow one to study, for example, correlations between interruptions and other factors like political party or age. They would also allow one to track
changes in the ratio of interruptions across time and to test whether such changes are influenced by factors like the topic under discussion and which party is in the government and which in opposition.

I will use the encoding of interruptions in the Austrian Parliamentary Records as an example.

### 2. Parliamentary Interruptions

Ilie (2005, 415) defines parliamentary discourse as “a process of joint real-life interaction that consists of interventions made by several participants involved in a dialogic exchange who are assigned particular institutional roles.” In most parliamentary debates, time on the floor is allocated by institutional rules and regulations rather than on a turn-by-turn basis. However, conversational interactions, such as interruptions, occur in parliamentary discourse (see Ilie 2001, 2005) and are an intriguing research topic.

Studies on parliamentary interruptions have been carried out by, among others, Wodak et al. (1990); Burkhardt (2004); Ilie (2005, 2010); Zima, Brône, and Feyaerts (2010); Stopfner (2013); and Truan (2016, 2017). Most of these studies are qualitative, applying close reading methods. Some of these studies focus on classifying parliamentary interruptions (e.g., Ilie 2005; Burkhardt 2004). According to Ilie (2005, 419), we can distinguish between verbal and nonverbal interruptions. Verbal interruptions are defined as “voiced intrusions in the current speaker's contribution” (Ilie 2005, 420), while nonverbal interruptions are, for example, applause or laughter. Interruptions can also be classified according to the initiator, whether an individual, such as a single member of parliament (MP), or a group, such as an entire party. Similarly, we can distinguish between institutionalized interruptions (e.g., a call to order) and noninstitutionalized interruptions (e.g., an interjection voiced by one MP). Finally, we can differentiate between positively and negatively connotated interruptions. Similar classifications can be found in Burkhardt (2004, 159ff). Regarding the classification, Ilie (2005, 419) states that “[n]one of the above parameters can adequately and exhaustively account on their own for the range of variation across interventions meant to interfere with the current speaker. Moreover, each of these parameters involve a number of shifting and overlapping variables.”
Even though interruptions are a general feature of parliamentary discourse, it has to be said that occurrence, type, and frequency depend on the parliamentary tradition and the rules of procedure (see Bevitori 2004, 87).

2.1 Recordings of Interruptions in Parliamentary Records

In order to study parliamentary discourse, and in particular parliamentary interruptions, researchers have to rely on the documentation of parliamentary debates, such as transcripts, shorthand records, and audio/video recordings. In most countries, parliamentary debates are recorded formally in parliamentary records and published. Winters (2017) states that “[p]arliamentary records form a unique longitudinal dataset about human behaviour: they can span hundreds of years, with periodical (often daily or weekly) accounts published according to a stable procedure. The nearly verbatim character of much of the transcription, albeit within certain conventions, also renders it closer to the spoken word than other sources with similar chronological coverage.”

Most parliamentary records, whether transcriptions or stenographic verbatim records, record not only the speeches of members of parliament, but also the interruptions. Typically, these interruptions are distinguished from the rest of the text by some typographical convention (e.g., italics, boldface, or indentation). I will present some examples taken from parliamentary records
of different countries in different languages. In the parliamentary records of the Italian Senate (see figure 1), for instance, the interruption is written in italics and enclosed in parentheses. The example shows a nonverbal interruption, namely applause (applausi).

Figure 1. Example of the recording of interruptions in the records of the Italian Parliament (XVIII, 186th sitting, January 30, 2020).

Senato della Repubblica – 12 – XVIII LEGISLATURA
186ª Seduta ASSEMBLEA - RESOCONTO STENOGRÁFICO 30 Gennaio 2020


Vorrei altresì sottolineare che la relazione del Parlamento è stata pronta: questo disegno di legge, che nasce anche per iniziativa dell’ex ministro dell’istruzione e della ricerca scientifica, lo portiamo a compimento in meno di un anno. C’è stata dunque una risposta molto pronta del Parlamento a un’esigenza del mondo scientifico, che penso accoglierà bene questo nostro lavoro. (Applausi dal Gruppo L-SP-Psd'Az).

AIROLA (M5S). Domando di parlare per dichiarazione di voto.
In the British Hansard, interruptions are marked by square brackets and italics. In the example in figure 2 we see a nonverbal interruption, namely laughter.

Figure 2. Example of the recording of interruptions in the records of the British Hansard (volume 670, no. 19, January 30, 2020).
In the following example of a German parliamentary record (figure 3), the interruptions are set in parentheses and indented, to contrast with the text of the regular speeches. In figure 3 we also see a combination of interruptions: a nonverbal interruption (Beifall, applause) and two verbal interruptions (by MP Michael Grosse-Brömer and MP Sören Bartol).

Figure 3. Example of the recording of interruptions in the German Parliamentary Records (19/143, January 30, 2020)³.

Präsidium Dr. Wolfgang Schäuble:
Nächst Redner ist der Kollege Dr. Joachim Pfeiffer, CDU/CSU:

(Beifall bei der CDU/CSU – Michael Grosse-Brömer [CDU/CSU]: Erzähle ihm noch mal, wie es hier funktioniert! In einfachen Sätzen! – Gegenruf des Abg. Sören Bartol [SPD]: Lohat sich gar nicht, darüber zu reden!)

Dr. Joachim Pfeiffer (CDU/CSU):
Herr Präsident! Meine sehr geehrten Damen und Herren! Liebe Kolleginnen und Kollegen! Die deutsche Glücksspirale dreht sich weiter, und zwar im elften Jahr. Es ist bereits vom Bundeswirtschaftsminister und auch vom Kollegen Linnemann angesprochen worden: Wir sind im elften Jahr des Wachstums, wenn auch moderat, und das gab es noch nie in der Geschichte der Bundesre-
In the Polish Parliamentary Records (figure 4), interruptions, both verbal and nonverbal, are also enclosed in parentheses. Nonverbal interruptions are set in italics, while verbal interruptions use a regular font style, with only the name of the MP or the collective set in italics.

Figure 4. Example of the recording of interruptions in the Polish Parliamentary Records (85th sitting, August 30, 2019).
Finally, the shorthand records of the Austrian National Council contain various types of interruptions in parentheses and italics (see figure 5).

Figure 5. Example of the recording of interruptions in the Austrian Parliamentary Records (XXIV, 10th sitting, January 21, 2009).

Figure 5, an example of an Austrian parliamentary record, illustrates a combination of verbal and nonverbal interruptions in parentheses and italics. Furthermore, the name of the MP voicing the interjection is set in bold type.

These examples show that interruptions recorded in the written records in these five countries range from terse descriptions like “Laughter” in the British Hansard or “Applausi” in the Italian records to more detailed descriptions, as in the Polish, German, and Austrian records, where the wordings of the verbal interruptions are also recorded.

In the next section we will closely examine these typographically marked interruptions in the Austrian Parliamentary Records. According to Klein (1982, 234), the jargon of the stenographers aptly refers to them as “Regiebemerkungen” (stage directions).
2.2 Austrian Use Case: Interruptions in the Austrian Parliamentary Records

As seen in figure 5, parliamentary interruptions in the Austrian Parliamentary Records are recorded in parentheses and italics. However, other transcriber comments are recorded in the same fashion as well. Overall, the following types of information can be found: verbal interruptions, nonverbal interruptions, descriptions of scenes or gestures, and descriptions of parliamentary procedures. Next, I will illustrate these types of information with examples.

**Verbal interruptions**

1.1 Individual verbal interruptions (by a specific person)
   Example: (Abg. Vilimsky: Volksabstimmung!)

1.2 Individual verbal interruptions, where the wording is not recorded by the stenographer
   Example: (Zwischenruf des Abg. Vilimsky.)

1.3 Collective verbal interruptions (by a party or group of MPs)
   Example: (Ruf bei den Grünen: Freudenau gibt’s nicht! – Ruf beim BZÖ: Was ist mit der Freudenau?)

1.4 Anonymous verbal interruptions
   Verbal interruptions where the wording is recorded by the stenographer but they are not attributed to an individual or collective
   Example: (Rufe: Warum?)

1.5 Verbal interruptions where the wording is not recorded by the stenographer and they are not attributed to a person or collective
   Example: (Zwischenrufe)

**Nonverbal interruptions**

2.1 Individual nonverbal interruptions (by a specific person)
   Example: (Abg. Dr. Khol schüttelt den Kopf.)

2.2 Collective nonverbal interruptions (by a party or a group of MPs)
   Example: (Beifall bei der SPÖ)
   Example: (Beifall bei den Freiheitlichen und bei Abgeordneten der ÖVP.)
2.3 Anonymous nonverbal interruptions not attributed to a person or collective (e.g., mentioning of a party)

Example: (Allgemeiner Beifall.)

Example: (Heiterkeit.)

Grouped interruptions

These are different types of interruptions grouped together in parentheses and separated by a dash.


Descriptions

Descriptions are of two types: descriptions of the scene, gestures, etc. as well as descriptions of the parliamentary procedure. The former can be related to interruptions, while the latter type is not an interruption.

4.1 Descriptions of the scene, gestures, etc.

This type of description can refer to a speaker or an interrupter and describes movements, gestures, or similar minutiae. The following examples are limited to descriptions of interrupters.

Example: (Abg. Mandak – auf die auf dem Rednerpult platzierte Tafel weisend –: Kann man das Taferl bitte wegnnehmen? – Abg. Kopf entfernt die Tafel vom Rednerpult.)

Example: (Abg. Grosz – in Richtung des sich zum Rednerpult begebenden Abg. Scheibner –: Bedank dich für die Neujahrswünsche!)

4.2 Description of parliamentary procedure

These descriptions are not interruptions as defined in section 2. They are included in this list for the sake of completeness.

Example: (Die Sitzung wird um 10.37 Uhr unterbrochen und um 10.44 Uhr wieder aufgenommen.)

One type of interruption that is mentioned in Ilie’s (2005) classification—the institutionalized interruptions, for example, a call to order (“Ordnungsruf”)—are not found as parenthetical comments in the Austrian Parliamentary Records. They are recorded as regular speech, as shown in figure 6.

Figure 6. Recording of a call to order in an Austrian Parliamentary Record (XXI, 100th sitting, April 17, 2002).

My goal in looking in detail at already existing classification of parliamentary records in the literature, at recordings of interruptions in different parliamentary records, and at the Austrian Parliamentary Records data was twofold: to elaborate a SKOS vocabulary to be used for the encoding (see section 4) and to recognize patterns in order to write rule-based scripts to identify and encode those interruptions automatically.

3. Available Parliamentary Corpora and Their Encodings

In section 2.1 we described the parliamentary records, usually available from the websites of the respective national parliaments in PDF or HTML format. In addition, there are parliamentary corpora containing collections of these parliamentary records in machine-readable form, enriched with metadata and various types of annotations.

A large number of parliamentary corpora is available (see Fišer, Lenardič, and Erjavec 2018, 1321) with a variety of different encodings. Looking at the available corpora, we can distinguish between three main types of encodings in parliamentary corpora:

1. custom or ad-hoc encodings, mostly in XML,
2. encodings in Akoma Ntoso, an XML-based format for parliamentary, legislative, and judiciary documents (Akoma Ntoso 2018), and
3. encodings in TEI (see Erjavec and Pančur 2019).

As we have seen in the previous examples of parliamentary records, interruptions are clearly marked and it is easy to detect and manually extract them (e.g., as demonstrated by Truan [2017]) by applying close-reading methods. But how is this realized in the digital sphere, and specifically in parliamentary corpora that are machine-readable? In the next section we will look at examples of parliamentary corpora and how they encode interruptions in order to develop computational methods for analysis. In this paper, we will not go into detail in the encodings in general, but will focus on how interruptions are encoded.
3.1 Corpora Encoded in Custom or Ad-hoc Encodings

Some examples of custom or ad-hoc encoded corpora are the CzechParl corpus (see Jakubíček and Kovář 2010), the beta version of the ParlAT Corpus (see Wissik and Pirker 2018), the collection of Swedish Parliamentary Data (see Rødven Eide 2019a), and the Knesset Corpus of transcriptions of Israeli parliament (see Itai and Wintner 2019). A comprehensive overview of corpora using custom encodings can be found in Erjavec and Pančur (2019).

Let us now have a closer look at the encodings of the Swedish Parliamentary Data used by Rødven Eide (2019a, 2019b) and how interruptions are encoded there. The data of the parliamentary speeches used in these encodings is downloaded directly from the website Riksdagens öppna data (Open Parliamentary Data)21 (see Rødven Eide 2019b) in different formats; in Example 1 an example in XML can be seen. The speeches are enriched with different types of metadata—for example, speaker <talar> and party <parti>—and the text of the speech is encoded with <anforandetext> and structured in paragraphs <p>. As shown in the example, the nonverbal interruption applause (Applåder) is encoded within the speech <anforandetext> as a paragraph <p>.


```xml
<anforande>
    <dok_hangar_id datatype="int32">5038419</dok_hangar_id>
    <dok_id datatype="string">H7092</dok_id>
    <dok_titel datatype="string">Protokoll 2019/20:2 Tisdagen den 10 september</dok_titel>
    <dok_rm datatype="string">2019/20</dok_rm>
    <dok_nummer datatype="int32">2</dok_nummer>
    <dok_datum datatype="datetime">2019-09-10 00:00:00</dok_datum>
    <avsnittsrubrik datatype="string">Riksmötets öppnande</avsnittsrubrik>
    <underrubrik datatype="string"/>
    <kammaraktivitet datatype="string"/>
    <anforande_id datatype="string">8bc861f9-4ee4-e911-9128-901b0e9b71a0</anforande_id>
    <anforande_nummer datatype="int32">2</anforande_nummer>
    <talare datatype="string">HANS MAJESTÅT KONUNGEN</talare>
    <parti datatype="string">HANS MAJESTÅT KONUNGEN</parti>
    
    <anforandetext datatype="string">Applåder</anforandetext>
    
    <p datatype="string">...</p>
```

Riksmötets öppnande

På den svenska politiska agendan stod vid denna tid bland annat försvarets inriktning och omfattning, Sveriges framtida energiförsörjning och miljöproblem såsom markförsurning, drivhuseffekten och påverkan på ozonlagret. I dag, tre årtionden senare, har utvecklingen i många avseenden gått framåt. Men på andra områden har utmaningarna snarare vuxit.

Herr talman! Ärade ledamöter! Riksdagsprotokollen berättar om vår tid för kommande generationer - om våra ideal och värderingar och om våra utmaningar och vår förmåga att möta dem. För 100 år sedan fattade Sveriges riksdag en rad avgörande beslut som tillsammans ledde fram till lika och allmän rösträtt. Jubileet påminner oss om vikten av att tänka och agera långsiktigt för vårt lands bästa.

Sveriges framtid formas inte enbart i denna kammare utan lika mycket av utvecklingen i vår omvärld. Men de beslut som fattas här inne påverkar ändå alla i vårt land, inte sällan generationer framöver. De är trådar i den väv som utgör vår framtid.


Härmed förklarar jag 2019/20 års riksmöte öppnat. (Applåder)

In the ParlAT beta version, for example, all interruptions were marked up with the XML element <comment> without distinguishing between verbal and nonverbal interruptions.
3.2 Corpora Encoded in Akoma Ntoso

As mentioned above, Akoma Ntoso is an XML-based format for parliamentary, legislative, and judiciary documents (Akoma Ntoso 2018) that is used for legal drafting processes and other workflows in various institutions. It is used only sporadically in the research community for building parliamentary corpora, mainly because the researchers who are building these corpora are not as familiar with the Akoma Ntoso standard as they are with, for example, TEI (see Erjavec and Pančur 2019). One research example is the study of Nigerian Senate debates by Lewis (2018).
In the example encoding from the Akoma Ntoso specification (Akoma Ntoso 2018) there are two different ways of encoding nonverbal interruptions: with the element `<remark type="scene description">` or with the element `<scene>`. For verbal interruptions no examples are given.

Example 2. Example encoding in Akoma Ntoso (Akoma Ntoso 2018).

```xml
<debate name="record">
  <meta>...</meta>
  <coverPage>...</coverPage>
  <preface>...</preface>
  <debateBody>
    <debateSection eId="dbsect_1" name="Preliminary">
      <rollCall eId="dbsect_1__rollCall_1">...</rollCall>
      <prayers eId="dbsect_1__prayers_1">...</prayers>
    </debateSection>
    <debateSection eId="dbsect_2" name="Actas">
      <speech eId="dbsect_2__speech_1" by="#Ascencio">
        <from>El señor ASCENCIO (Presidente).-</from>
        <p>El señor Secretario dará lectura a los pareos.
          <remark type="sceneDescription">(Aplausos)</remark></p>
        ......
      </speech>
      <narrative eId="dbsect_2__narrative_1">-Con posterioridad, la Sala se pronunció sobre el proyecto en los siguientes términos:</narrative>
      <summary eId="dbsect_2__summary_1">
        <outcome refersTo="#seAbstiene">-Se abstuvieron los diputados señores:
          <vote by="#persona">
            <person eId="p3517" refersTo="#GalileaVidaurreJoseAntonio">Galilea Vidaurre José Antonio;</person>
          </vote>
        </outcome>
      </summary>
      <scene eId="dbsect_2__scene_1">-Aplausos.</scene>
    </debateSection>
  </debateBody>
</debate>
```
3.3 Corpora Encoded in TEI

There are several parliamentary corpora that are encoded in TEI (Erjavec and Pančur 2019) using one or more TEI modules. Within the Parla-CLARIN initiative a set of recommendations for harmonizing the TEI encoding of parliamentary corpora has been proposed (Erjavec and Pančur 2019). The most popular modules are the performance text module (drama) and the transcription of speech module (spoken). I will now take a closer look at the encoding of interruptions in some of these corpora.

For the automatic encoding of the GermaParl (Blätte and Blessing 2018), the corpus of the German parliamentary records, the TEI module for performance text was used. The first version of the SlovParl Corpus, the corpus of Slovenian Parliamentary Records, was also encoded using this module (see Pančur 2016), while the Danish Parliamentary Corpus uses elements from the transcriptions of speech module (see Haltrup Hansen and Navaretta 2019). The performance text module is intended for the encoding of printed dramatic texts, screenplays, or written transcriptions of other types of performances according to the TEI P5 Guidelines (see TEI Consortium 2019). Using it for parliamentary records is not a completely inappropriate choice, because there are similarities in interruption patterns in parliamentary discourse and drama dialogue, as discussed by Ilie (2005) and by Klein (1982, 234). As noted above, Klein even points to stenographers’ jargon that calls the notes of interruption “Regiebemerkungen” (stage directions).

As an example for a corpus using the performance text module, we will examine the GermaParl corpus. In this corpus the following information is annotated: the speaker’s name, the party affiliation, and whether an utterance is a speech or an interjection (see Blätte and Blessing 2018, 812). Individual utterances in speeches are annotated as <sp> with paragraphs as <p>. Interjections are encoded as <stage type="interjection"/> within speeches. However, as can be seen in
example 3, the encoders do not distinguish between nonverbal interruptions such as “(Beifall bei der FDP und der CDU/CSU)” and verbal interruptions such as “(Hans Büttner [Ingolstadt] [SPD]: Immer schon!)” which makes computer-based analysis of interruptions more difficult.

Example 3. Example encoding of interruptions in the GermaParl Corpus (BT, October 17, 2002).

```
<sp who="Jörg van Essen" parliamentary_group="FDP" role="mp" position="NA" name="Jörg van Essen">
  <speaker>Jörg van Essen (FDP):</speaker>
  <p>Herr Präsident! Ich gratuliere Ihnen auch im Namen meiner Fraktion sehr herzlich zu Ihrer Wiederwahl.

  Wir haben mit sehr großer Freude gehört, dass Sie die Wahrung der Interessen des ganzen Bundestages und aller Abgeordneten in den Mittelpunkt Ihrer zweiten Wahlzeit stellen wollen. </p>
  <stage type="interjection">(Hans Büttner [Ingolstadt] [SPD]: Immer schon!)</stage>

  <p>Sie haben dabei unsere volle Unterstützung.</p>
  <stage type="interjection">(Beifall bei der FDP und der CDU/CSU)</stage>
  <p>Die FDP-Bundestagsfraktion unterstützt den Antrag der CDU/CSU-Fraktion. Wer sich das Wahlergebnis anschaut, stellt fest, dass es wenige Tausend Stimmen Unterschied zwischen beiden großen Parteien gibt.

  Die SPD stellt den Präsidenten, die SPD stellt eine Vizepräsidentin - also zwei Personen im Präsidium.

  Dann ist es doch nur eine Frage der Fairness, dass auch die zweite große Fraktion mit zwei Personen im Präsidium des Deutschen Bundestages vertreten ist.</p>

  <stage type="interjection">(Beifall bei der FDP und der CDU/CSU)</stage>
  <p>Diese schlichte Fairness, diese Gerechtigkeit</p>
  <stage type="interjection">(Wilhelm Schmidt [Salzgitter] [SPD]: Vor vier Jahren waren Sie anderer Meinung)</stage>

  <p>sollte unsere Zusammenarbeit bestimmen.

  Das Argument, das der Kollege Schmidt gebracht hat, dass sich die Mehrheit auch im Präsidium widerspiegeln müsse, trifft gar nicht zu.

  Die letzte Legislaturperiode - da hatten wir auf beiden Seiten jeweils drei Vertreter - hat gezeigt, dass eine Zusammenarbeit möglich ist.

  Dies hat zu keinerlei Nachteilen geführt. Ich bin sicher, dass es auch hier wieder so wäre.</p>
```
Von daher wird von der FDP-Bundestagsfraktion mit Nachdruck unterstützt, dass es einen zweiten Vizepräsidenten oder eine Vizepräsidentin für die zweitgrößte Fraktion im Deutschen Bundestag gibt.

(Beifall bei der FDP und der CDU/CSU - Hans Büttner [Ingolstadt] [SPD]: Heuchelei ist das!)

The other TEI module that is often used for encoding parliamentary data is the transcription of speech module. This is not surprising since parliamentary records, especially the parts covering plenary debate, are close to the spoken word, despite being edited. For example, in the Austrian Parliamentary Records, voiced pauses such as “äh” or grammatical errors are removed, dialectal elements are changed into standard elements, and at the end the MPs also have the right to ask for stylistic changes (see Stopfner 2013). In the UK, the reporters use, for example, the Hansard style guide when editing the records (see Shaw 2018). Among others, SlovParl 2.0 (Pančur, Šorn, and Erjavec 2018), the TAPS-fr corpus, the corpus of French parliamentary debates (see Diwersy, Frontini, and Luxardo 2018), and de-parl (Parliamentary Debates on Europe at the Deutscher Bundestag [1998–2005]) (see Truan 2019a) use elements of the transcription of speech encoding. In the manually annotated de-parl corpus the speeches and verbal interruptions are encoded as utterances with the element <u>. Nonverbal interruptions such as applause are annotated with <incident> and other comments by the stenographers with <note> (see Truan 2019b, 10). In example 4, utterances and incidences can be identified. However, without having the original records there is no way to categorize the utterance

Sie sind doch in der Regierung!

as a verbal interruption, an interjection. There is no attribute with a value indicating that in the original transcript this <u> was an interruption, typographically marked as intended text within brackets. As a consequence, it is impossible, relying only on the encoded corpus, to distinguish between speech and verbal interruptions, and therefore to extract, for example, only the verbal
interruptions. Furthermore, there is no markup for interrupted speech, such as speech by the same speaker that continues after the interruption, which in the original is implied with the dash “–” at the beginning of the sentence (see figure 7).

Example 4. Example from the de-parl encoding (BT, December 10, 1998)\textsuperscript{24}.

\begin{verbatim}
<u who="#FISCHER">Herr Präsident! Meine Damen und Herren! Herr Kollege Haussmann, wenn
Sie meinen, daß das europäische Anliegen besser durch solche Vorträge voran
gebracht werden kann, wie Sie ihn hier gehalten haben, und wenn Sie dem Kollegen Struck
vorwerfen, daß er eine ruhigere Gangart gewählt hat, dann kann ich Ihnen nur
folgendes sagen; ich habe Ihnen da sehr sorgfältig zugehört.</u>

<incident>
<desc>Widerspruch bei der CDU/CSU und der F.D.P.</desc>
</incident>

<u who="#HAUSSMANN">Das ist ja peinlich!</u>

<u who="#FISCHER">Herr Kollege Haussmann, in der Zeit, in der ich hier im Plenum
war, <incident>
<desc>Lachen bei der F.D.P. sowie bei Abgeordneten der CDU/CSU</desc>
</incident>habe ich Ihnen sehr sorgfältig zugehört.</u>

<u who="#HAUSSMANN">Ein bißchen billig!</u>

<u who="#FISCHER">Das ist überhaupt nicht billig.</u>

<u who="#HAUSSMANN">Doch! Aber Sie sind nicht kollegial!</u>
\end{verbatim}
The Parla-CLARIN recommendations are “intended for the encoding of corpora of parliamentary proceedings, regardless of the language or country of origin, for the purposes of scholarly investigations, be they from the field of linguistics, political science, history or other humanities and social sciences disciplines. The recommendations are, in principle, not meant as the primary storage format of parliamentary proceedings, such as kept by governmental offices, for which Akoma Ntoso might be preferred” (Erjavec and Pančur 2022). In Parla-CLARIN the transcription of speech module (spoken) is used in addition to other modules like TEI Corpus (corpus) or TEI person (namesdates, see Erjavec and Pančur 2019). In the Parla-CLARIN guidelines there are specific sections dealing with interrupted utterances and incidents. The recommendations state that although an “interruption might be ... encoded as a <note>... , it is more precisely encoded as a separate utterance” (Erjavec and Pančur 2022). However, the recommendations do not suggest any attributes to indicate that these utterances are verbal interruptions or interjections. Once they are encoded as <u> they are not distinguishable from other speech encoded as <u> unless they are
splitting utterances. In this case the guidelines suggest using the @next attribute on the first part of the split utterance and the @prev attribute on the second part. However, not all interjections or interruptions split a speech; they can also occur at either end, where there are no means to distinguish them from other utterances.

As for nonverbal interruptions, Parla-CLARIN suggests using <note> for transcribers’ comments in general. For more specific purposes, the following elements from the transcription of speech module can be used:

- `<vocal>` marks any vocalized but not necessarily lexical phenomenon, e.g., laughter, sounds of (dis)agreement from the benches etc.
- `<kinesic>` marks any communicative phenomenon, not necessarily vocalized, for example a gesture, frown, etc.
- `<incident>` marks any phenomenon or occurrence, not necessarily vocalized or communicative, for example incidental noises or other events affecting communication.

(Erjavec and Pančur 2022 sec. 5.2: Transcriber comments)

4. Encoding Interruptions: Experiences from the Austrian Parliamentary Records Corpus

As described in section 2.2, several types of interruptions can be found in the Austrian Parliamentary Records in the form of transcribers’ notes. In this section, we will describe how these notes can be encoded in order to facilitate quantitative computational analysis and computer-aided studies of parliamentary interruptions.

As a general rule, we followed the Parla-CLARIN recommendations (Erjavec and Pančur 2022), adding more detailed encodings as necessary, especially through attributes. As we have seen in section 2.1, the Austrian records provide some of the most detailed descriptions of interruptions. Designing an encoding scheme with the Austrian Parliamentary Records in mind, we will offer
an encoding suggestion that is fine-grained enough to deal with their level of detail regarding interruptions. That same scheme, because it relies upon the Parla-CLARIN recommendations, can be used to encode records containing less-detailed descriptions.

To organize these attributes, further described below, we used SKOS vocabularies. We created two concept schemes: one for utterances with the prefix ucat: (which stands for “utterance categories”) and one for notes with the prefix ncat: (for “note categories”). The definition of the prefix is included in a <prefixDef> element in the <encodingDesc> and points to the SKOS vocabulary. The values of @ana are defined in a scope note. In the following sections we will describe the encodings for verbal interruptions and nonverbal interruptions.

4.1 Encoding Verbal Interruptions

We encoded verbal interruptions where the wording is recorded as utterances <u>; we encoded verbal interruptions where the wording is not recorded like nonverbal interruptions (see section 4.2 and example 6), as notes. In order to distinguish verbal interruptions encoded as utterances from the authorized speeches, which are also encoded as utterances, we used the attribute @ana with the value “ucat:unauthorized” for interjections (see example 5) and the value “ucat:regular” for authorized speeches. Interjections frequently interrupt speeches and, therefore, we used the attributes @prev and @next to connect parts of interrupted speeches. This practice is also suggested by the Parla-CLARIN recommendations. We also transformed verbal interruptions by a collective into utterances encoded with <u>. The earlier example (Ruf bei den Grünen: Freudenau gibt’s nicht! – Ruf beim BZÖ: Was ist mit der Freudenau?) was split up into two utterances. Usually the @who attribute is associated with a specific individual. Since the interruption at hand is by a collective, no single speaker is specified. To handle such cases, we introduced <personGrp>s for the collectives in each file header, to which the @who attribute can point (e.g., "#SPEAKER_COLLECTIVE_Grünen", "#SPEAKER_COLLECTIVE_BZÖ", and analogously for any other party). In our encoding proposal, we not only encode the utterance itself, as for example Truan (2016, 2019a) did in de-parl, but we also encode the associated description in the parliamentary records. In the example above there are the phrases “Ruf bei den Grünen:” and “Ruf beim BZÖ,” which denote interjections from the Green party and from the BZÖ party. These descriptions are encoded as <note> with @type value "comment" and @ana value
Encoding Interruptions in Parliamentary Data: From Applause to Interjections and Laughter

"ncat:interjection_speaker_collective". An example of the encoding can be seen in example 5. For interjections where the wording is recorded by the stenographer but which are anonymous, no @who attribute is assigned.

Example 5. Encoding of verbal collective interruptions.

<note ana="ncat:interjection_speaker_collective" type="comment"> Ruf bei den Grünen: </note>

<u ana="ucat:unauthorized" xml:id="d7e8727" who="#SPEAKER_COLLECTIVE_GRÜNEN"> Freudenau gibt’s nicht! </u></seg>

<note ana="ncat:interjection_speaker_collective" type="comment"> Ruf beim BZÖ: </note>

<u ana="ucat:unauthorized" xml:id="d7e8735" who="#SPEAKER_COLLECTIVE_BZÖ"> Was ist mit der Freudenau? </u></seg>

4.2 Encoding Nonverbal Interruptions

As described in the Parla-CLARIN recommendations, transcribers’ comments are encoded as <note>. We followed these recommendations. Since transcribers’ comments may cover more than just interruptions (see section 2.2) and also because other information is commonly encoded as <note> we used the following values for @type: "speaker", "comment", and "time" (see also Erjavec and Pančur 2022). The notes of @type="comment" were further subcategorized with the attribute @ana, which can have the following values to describe a nonverbal interruption: "ncat:applause", "ncat:bell", or "ncat:laughter" (see example 6). Moreover, descriptions related to verbal or nonverbal interruptions can be encoded as <note> of @type value "comment" with the following values for @ana: "ncat:direction" to describe if the interrupter is facing toward a specific person or party group; "ncat:nonverbal_hands" to denote that the interrupter is holding something up like a sign, document, newspaper article, or picture; this encoding may be further combined with "ncat:inscription" in cases where the sign’s content is reported; and, finally, "ncat:nonverbal" for other nonverbal phenomena that are not covered by the values specified above.
Furthermore, as mentioned in section 4.1, verbal interruptions where the wording is not recorded are also encoded as <note> using an @ana value of "ncat:interjection" (see example 7). Here are two examples of encodings for nonverbal interruptions:

Example 6. Example of the encoding of a nonverbal interruption.

<note anana="ncat:applause" type="comment"> (Beifall beim BZÖ.) </note>

Example 7. Example of the encoding of a verbal comment, where the wording is not recorded.

<note anana="ncat:interjection" type="comment">(weitere Zwischenrufe bei der ÖVP) </note>

It might seem inconsistent that we use both @type and @ana and do not choose one or the other. There are two main reasons for this. First, @type is also used in the Parla-CLARIN recommendations and we wanted to be as compatible as possible with them in order to do comparative studies with other corpora that are encoded according to the Parla-CLARIN recommendations. Second, the values for @type, like "speaker", "comment", and "time", are less interpretative and can be inferred from the text, whereas the values we used for @ana are more interpretative. Therefore we use both @type and @ana.

5. Conclusions

Parliamentary data is a comprehensive and unique resource for many different fields in the humanities and social sciences. In this paper we have shown that research on parliamentary interruptions is one of the areas that can benefit from the growing availability of machine-readable and annotated parliamentary corpora especially regarding computer-based quantitative analysis.

In this paper, I have shown that although interruptions are clearly marked in the original parliamentary records, and are easily recognizable using close reading methods, they are, nevertheless, not always encoded unambiguously in machine-readable parliamentary corpora and, therefore, are not easily extractable. I have discussed how the Parla-CLARIN recommendations provide a good basis for encoding parliamentary corpora and that they already include guidelines on how to encode interjections and nonverbal interruptions. However, they are only a basic framework; more detailed encoding of verbal and nonverbal interruptions are needed to enable the analysis of parliamentary interruptions. In this paper, I have proposed recommendations for
Encoding interjections to distinguish them from authorized speech and to differentiate between types of nonverbal interruptions. My examples were drawn from the encoding of the interruptions in the Austrian Parliamentary Records. The recommendations are customizable, and the chosen values defined in the SKOS vocabulary will depend on the specific data being encoded. In summary, I argue that a more detailed encoding of parliamentary interruptions, as explained in this paper, is needed in order to make verbal and nonverbal interruptions identifiable and extractable and to facilitate computer-based quantitative analysis.

BIBLIOGRAPHY


NOTES


AUTHOR

TANJA WISSIK

Tanja Wissik is a senior scientist at the ACDH-CH (Austrian Centre for Digital Humanities and Cultural Heritage) of the Austrian Academy of Sciences and teaches at the University of Graz and the University of Vienna. She works on numerous national and international research projects in the fields of language resources and digital humanities. Her research interests include digital language resources and technologies, especially terminological resources and corpora, variation and language change, workflow research, and the teaching of DH technologies and methods.