



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DIPARTIMENTO DI INTERPRETAZIONE E TRADUZIONE - DIT

CORSO DI LAUREA MAGISTRALE IN
INTERPRETAZIONE

**WAS IST DER REDE WERT? UN'ANALISI
CORPUS-BASED DELL'USO DEL LINGUAGGIO
DI GENERE NEL TEDESCO PARLATO**

Tesi di laurea magistrale in Interpretazione dal Tedesco in Italiano

Relatrice

Prof.ssa Gabriele Dorothe Mack

Presentata da

Dalila D'Andrea

Correlatrice

Prof.ssa Doris Anita Höhmann

Sessione marzo 2026

Anno Accademico 2024/2025

Indice

Indice	2
Lista delle figure	5
Lista delle tabelle	5
Introduzione.....	6
Abstract.....	8
Zusammenfassung.....	9
PE3IOME.....	10
CAPITOLO 1 – GENERE E LINGUAGGIO	11
1.1 Genere e Linguaggio	11
1.1.2 La binarietà di genere e la terza opzione	13
1.2 Il maschile generico e il <i>Male As Norm</i>.....	15
1.3 La linguistica femminista.....	16
1.3.1 Luise Pusch – Das Deutsche als Männersprache	19
1.4 Lingua, pensiero e marginalizzazione	21
1.5 Reazioni e sviluppi: le prime linee guida.....	23
1.6 Il linguaggio di genere in tedesco oggi.....	26
1.7 Le principali soluzioni per l'impiego di un linguaggio inclusivo in tedesco	30
1.7.1 Il colpo di glottide.....	36
CAPITOLO 2 – I CORPORA.....	37
2.1 La linguistica dei corpora	37
2.2 Linguistica dei corpora e linguaggio di genere	39
2.3 La creazione di un corpus orale	41
2.3.1 Il 'written language bias' e il divario nella ricerca su tedesco scritto e orale	42
2.3.2 Caratteristiche della comunicazione orale	44
2.4 La creazione del corpus <i>Alles Gesagt?</i>	45
2.4.1 La scelta del materiale: Luise Pusch e il podcast <i>Alles Gesagt?</i>	45

2.4.2 La fase di trascrizione	48
2.4.3 La trascrizione supportata da riconoscimento automatico del parlato	49
2.4.4 La trascrizione del parlato dialogico supportata da tecnologie IA	51
- Obiettivi del progetto e strumenti impiegati	51
- L'organizzazione del lavoro tramite <i>pipeline</i>	53
- Limiti e potenzialità	54
2.4.5 Annotazione e interrogazione del corpus: Sketch Engine e GermanRFTagger	55
2.5 Il corpus <i>Politische Reden</i> (1982-2020) del DWDS	56
2.5.1 Interrogazione del corpus <i>Politische Reden</i> (1982-2020)	58
2.6 I due corpora a confronto	58
CAPITOLO 3 – ANALISI DEI CORPORA	59
3.1 Analisi del corpus <i>Alles Gesagt?</i>	59
3.1.1 Sull'utilizzo di <i>man</i> e delle sue alternative	59
3.1.2 Sull'utilizzo di sdoppiamento e colpo di glottide	63
3.1.3 Sull'uso di aggettivi e participi sostantivati nel corpus <i>Alles Gesagt?</i>	65
- Participi presenti	65
- Participi passati	66
- Aggettivi sostantivati	67
3.1.4 Sull'uso di collettivi ed epiceni nel corpus <i>Alles Gesagt?</i>	68
- Collettivi	68
- Epiceni	69
3.1.5 Alcuni pronomi indefiniti e la loro ripresa anaforica	71
3.1.6 Le strategie di genere utilizzate a confronto	71
3.2 Analisi del corpus <i>Politische Reden</i> (1982-2020)	74
3.2.1 Alternative a <i>man</i>	74
3.2.2 Sdoppiamento in forma contratta	74
3.2.3 Analisi per coppie di <i>Personenbezeichnungen</i>	76

3.3 Discussione dei risultati	80
Conclusioni.....	83
Bibliografia	85
Sitografia	93
Appendice: Un nuovo approccio alla trascrizione automatica del parlato dialogico	96
La trascrizione e <i>diarizzazione</i> del parlato dialogico supportata da tecnologie di IA	96
Introduzione – Il progetto.....	96
<i>Whisper</i> e <i>PyAnnote</i>, punti di forza e criticità.....	97
Scelte tecnologiche.....	98
<i>PyAnnote Audio 4.0</i> con modello <i>speaker-diarization-community-1</i>	98
Modalità <i>Exclusive Mode</i> in <i>PyAnnote</i>	99
<i>Whisper</i> di OpenAI	99
Pipeline a blocchi per audio lunghi.....	100
Mappatura speaker semi-automatica e automatica	100
Filtraggio a due stadi (merge + durata).....	100
Blocchi audio ottimali per <i>Whisper</i>	101
Revisione delle trascrizioni tramite uno strumento per il sottotitolaggio	101
Approccio per fasi	101
Le fasi nel dettaglio.....	101
Output finali	103
Limitazioni note.....	103
Ringraziamenti	104

Lista delle figure

Figura 1 – Frequenza di linguaggio inclusivo di genere dal 2000 al 2021 nel corpus DeReKo (da Waldendorf, 2024: 363).....	27
Figura 3 – Tipi di sdoppiamento in forma contratta nel corpus	75
Figura 4 - Uso di <i>splitting</i> , trattino basso e <i>Binnen-I</i> all'interno del corpus nel corso del tempo	75
Figura 5 – Forma maschile e forme alternative in ottica diacronica per la coppia <i>Bürger</i> + <i>Bürgerin</i> (frequenze assolute)	77
Figura 6 – Forma maschile e forme alternative in ottica diacronica per la coppia <i>Mitarbeiterin</i> + <i>Mitarbeiter</i> (frequenze assolute)	78
Figura 7 – Forma maschile e forme alternative in ottica diacronica per la coppia <i>Wählerin</i> + <i>Wähler</i> (frequenze assolute).....	79

Lista delle tabelle

Tabella 1 – Proposte di ordine terminologico.....	31
Tabella 2 – Proposte di ordine sintattico	34
Tabella 3 – Principali caratteristiche dei due corpora	58
Tabella 4 – Uso di <i>frau</i> e <i>man</i> nel subcorpus <i>Pusch</i> e nel subcorpus <i>Intervistatori</i>	60
Tabella 5 – Frequenze delle varie forme nel dettaglio	61
Tabella 6 – Contesti d'uso dei pronomi <i>frau</i> e <i>eine</i> nel subcorpus <i>Pusch</i>	63
Tabella 7 – Sdoppiamento per esteso e colpo di glottide nei due subcorpora	64
Tabella 8 – Participi sostantivati nei due subcorpora	67
Tabella 10 – Collettivi più frequenti nei due subcorpora	68
Tabella 11 – Epiceni maggiormente frequenti nei due subcorpora	70
Tabella 12 – Le strategie di genere analizzate nel corpus <i>Alles Gesagt?</i>	72

Introduzione

Negli ultimi decenni il dibattito intorno all'impiego di un linguaggio rispettoso ed inclusivo nei confronti di tutti i generi si è intensificato fino a diventare uno dei temi più dibattuti e controversi del discorso pubblico in area germanofona. La critica linguistica femminista a partire dagli anni Settanta ha messo in luce come determinate strutture linguistiche contribuiscano alla costruzione e al mantenimento di gerarchie che privilegiano un genere rispetto agli altri. Ne consegue che l'impiego del linguaggio di genere non rappresenta una mera questione stilistica o formale, ma si configura piuttosto come un processo linguistico che si intreccia con dimensioni profonde del rapporto tra lingua, pensiero e società, e diventa quindi uno strumento chiave per il raggiungimento della parità di genere.

Alla luce della crescente attenzione che il tema dell'inclusività ha assunto a livello nazionale ed internazionale, la necessità di individuare strategie e soluzioni per esprimersi in maniera al contempo corretta e inclusiva assume rilevanza centrale per linguisti/e, interpreti, traduttori/trici, ma anche attivisti/e, politici e comunicatori/trici di ogni genere.

Il presente elaborato si propone di analizzare, in chiave *corpus-based*, alcuni aspetti legati all'impiego del linguaggio di genere nel tedesco parlato. L'obiettivo è esplorare diverse sfaccettature del fenomeno, osservandone la realizzazione in contesti comunicativi concreti, mettendo in relazione dimensione qualitativa e quantitativa e sfruttando le nuove possibilità offerte dall'avvento delle tecnologie di IA nell'ambito della trascrizione del parlato. Il forte focus sulla lingua parlata risponde alla necessità di colmare un divario nella ricerca, tradizionalmente orientata verso la lingua scritta. Inoltre, questo approccio permette di esplorare se e come è possibile applicare il linguaggio di genere non solo al linguaggio politico o burocratico-amministrativo scritto, ma anche al parlato quotidiano, offrendo spunti preziosi per chiunque voglia "imparare" a esprimersi in modo al contempo naturale e rispettoso, compresi gli/le interpreti che desiderano implementare e automatizzare l'uso del linguaggio di genere nella pratica interpretativa di tutti i giorni.

Per l'analisi sono stati impiegati due corpora: uno creato ad hoc (*Alles Gesagt?*) costituito dalla trascrizione di un episodio di podcast in cui viene intervistata la linguista e femminista tedesca Luise Pusch, e uno pubblicamente disponibile sul sito del DWDS (*Politische Reden (1982-2020)*) e composto da discorsi tenuti da figure politiche di spicco dell'area germanofona tra il 1982 e il 2020.

Il lavoro di tesi si divide in tre capitoli. Il primo capitolo offre il quadro teorico di riferimento, introducendo alcuni concetti chiave del legame tra genere e linguaggio, illustrando le principali soluzioni proposte dalle linee guida per il linguaggio di genere in tedesco e tracciando una panoramica dell'evoluzione storica della linguistica femminista tedesca e del linguaggio di genere in tedesco dagli anni Settanta ad oggi.

Il secondo capitolo descrive i metodi e i materiali adottati: viene offerta una panoramica sulla linguistica dei corpora e sulla sua applicazione allo studio del linguaggio di genere e vengono illustrate le caratteristiche e le modalità di creazione dei corpora utilizzati, di cui vengono evidenziati limiti e potenzialità.

Il terzo capitolo è dedicato all'analisi dei fenomeni e delle strategie linguistiche selezionate e alla discussione dei risultati. L'analisi si articola in due parti: l'analisi del corpus *Alles Gesagt?* verte sulle soluzioni di genere presenti nel parlato di Pusch e dei suoi due intervistatori, analizzate in chiave qualitativa; l'analisi del corpus *Politische Reden (1982-2020)*, più marcatamente quantitativa e diacronica, permette di osservare lo sviluppo di diverse varianti di linguaggio inclusivo nel tempo sulla base di alcuni casi esemplari.

La presente tesi è da intendersi come uno studio pilota esplorativo che mira a contribuire alla riflessione sull'uso del linguaggio di genere nel tedesco parlato, offrendo al contempo spunti di interesse per la formazione e la pratica professionale in ambito traduttivo e interpretativo e per possibili studi futuri.

Abstract

The object of this Master thesis is a *corpus-based* exploration of the use of Gender Inclusive Language (GIL) in the spoken German language, conducted on two *spoken language* corpora. One was created *ad hoc* (*Alles Gesagt?*) by transcribing a podcast episode where the German linguist and feminist Luise Pusch is interviewed, and one publicly available on the DWDS website (*Politische Reden (1982-2020)*). The research made use of the state-of-the-art AI-powered tools *WhisperAI* and *pyAnnote* for the transcription process, while Sketch Engine and the dedicated DWDS online interface have been used to interrogate the corpora. The thesis is divided into three chapters: the first chapter focuses on the link between language and gender as well as on the evolution of GIL in the German language; the second chapter is dedicated to corpus linguistics and illustrates the characteristics of the selected corpora; the last chapter contains the analysis of both corpora. The analysis of the smaller, *ad hoc* corpus is mainly qualitative and sheds light on how different forms of GIL can be used in the spoken language. It highlights a significant use of strategies such as the glottal stop, the double gender marking and gender-neutral words such as epicene and collective nouns. The quantitative diachronic analysis carried out on the second, larger corpus, on the other hand, allows to observe the evolution of different GIL alternatives over a forty-year period (1982-2020), indicating a significant shift towards gender-balanced structures in the German political language over the years, with binary inclusive forms such as the double gender marking increasing their presence as an alternative to the traditional generic masculine, while non-binary inclusive forms have slowly surfaced over the last twenty years. To conclude, the results are summarized, and possible future developments discussed.

Zusammenfassung

Ziel der vorliegenden Masterarbeit ist es, die Verwendung geschlechtergerechter Sprache im gesprochenen Deutsch korpuslinguistisch zu untersuchen. Zu diesem Zweck wurden zwei Korpora der gesprochenen Sprache verwendet: ein selbsterstelltes Korpus (*Alles Gesagt?*), das auf der Transkription einer Podcast-Episode mit der Linguistin und Feministin Luise Pusch basiert, sowie das öffentlich zugängliche Korpus *Politische Reden (1982-2020)* des DWDS. Zur Optimierung des Transkriptionsprozesses wurden moderne KI-Tools (*WhisperAI* und *pyAnnote*) verwendet, während für die Untersuchung der Korpora Sketch Engine und die DWDS-Korpussuche Verwendung fanden.

Die Arbeit gliedert sich in drei Kapitel: Das erste Kapitel behandelt den Zusammenhang zwischen Sprache und Geschlecht sowie die Entwicklung geschlechtergerechter Sprache im Deutschen; das zweite Kapitel widmet sich der Korpuslinguistik und veranschaulicht die Merkmale der untersuchten Sprachstrukturen; das dritte und letzte Kapitel enthält das Kernstück der vorliegenden Arbeit, den geleisteten empirischen Beitrag zur Untersuchung der geschlechtergerechten Sprache im Deutschen.

Die Analyse des kleineren selbsterstellten Korpus ist hauptsächlich qualitativ und beleuchtet, wie verschiedene geschlechtergerechte Formen in der gesprochenen Sprache vorkommen, wobei der signifikante Einsatz von Strategien wie dem Glottisschlag, der Beidnennung sowie geschlechtsneutralen Begriffen (etwa Epikoina und Sammelbegriffe) hervorgehoben wird.

Die quantitative diachron ausgerichtete Analyse des größeren Korpus ermöglicht hingegen die Beobachtung der Entwicklung verschiedener geschlechtergerechten Alternativen über einen Zeitraum von fast vierzig Jahren (1982–2020) und deutet auf eine deutliche Verschiebung zu geschlechtergerechten Strukturen in der deutschen politischen Sprache hin, wobei binäre inklusive Formen wie die Beidnennung zunehmend das traditionelle generische Maskulinum zu ersetzen scheinen. Auch nicht-binäre inklusive Formen betreten in den letzten zwanzig Jahren verstärkt die sprachliche Bühne. Abschließend werden die Ergebnisse zusammengefasst und mögliche zukünftige Entwicklungen diskutiert.

РЕЗЮМЕ

Предметом данной магистерской диссертации является основанное на корпусе исследование использования гендерно-нейтрального языка в устной речи на немецком языке, проведенное на двух устных корпусах. Один из них был создан специально (*Alles Gesagt?*) путём транскрибирования эпизода подкаста, в котором интервьюируется немецкий лингвистка и феминистка Луиза Пуш, а другой доступен на сайте DWDS (*Politische Reden (1982-2020)*). В исследовании использовались современные инструменты искусственного интеллекта *WhisperAI* и *pyAnnote* для транскрибирования, а для анализа корпусов были использованы *Sketch Engine* и специальный онлайн-интерфейс DWDS.

Диссертация состоит из трех глав: первая глава посвящена связи между языком и гендером, а также эволюции гендерно-нейтрального языка в немецком языке; вторая глава посвящена корпусной лингвистике и иллюстрирует характеристики выбранных корпусов; последняя глава содержит анализ обоих корпусов.

Квалитативный анализ меньшего, специально созданного корпуса проливает свет на то, как различные формы гендерно-нейтрального языка могут использоваться в устной речи. В нем подчеркивается значительное использование таких стратегий, как гортанная смычка (*Glottisschlag*), упоминание обоих родов (*Beidnennung*) и гендерно-нейтральные слова.

С другой стороны, диахронический квантитативный анализ второго, более крупного корпуса позволяет наблюдать эволюцию различных гендерно-нейтральных альтернатив на протяжении сорока лет (1982-2020), указывая на рост гендерно-нейтральных форм в немецком политическом языке. При этом бинарные инклюзивные формы все чаще используются в качестве альтернативы традиционному общему мужскому роду, а небинарные инклюзивные формы постепенно появились в последние двадцать лет. В заключение подводятся итоги и обсуждаются возможные будущие исследования.

CAPITOLO 1 – GENERE E LINGUAGGIO

Il presente capitolo offre una panoramica sul legame tra genere e linguaggio, con un focus particolare sul panorama di lingua tedesca. Innanzitutto, vengono introdotti alcuni concetti chiave come *Genus*, *Sexus* e *Gender*, e viene aperta una parentesi informativa sulla binarietà (e non binarietà) di genere (1.1). Segue una sezione dedicata all'uso del cosiddetto maschile generico e al concetto del *Male As Norm* (1.2). Successivamente, viene ripercorsa a grandi linee la storia della linguistica femminista tedesca, con un'attenzione particolare al libro *Das Deutsche als Männersprache* di Luise Pusch (1.3). Viene poi dedicata una sezione al legame tra linguaggio, pensiero e marginalizzazione (1.4). Successivamente, vengono illustrati gli sviluppi nel campo del linguaggio di genere in tedesco dagli anni Settanta ad oggi (1.5 e 1.6). Infine, vengono presentate due tabelle riassuntive che illustrano le principali soluzioni proposte dalle linee guida per il linguaggio di genere in tedesco (1.7) e viene dedicata una sezione al fenomeno del colpo di glottide.

1.1 Genere e Linguaggio

Quello che comunemente in Germania viene chiamato *Gendern*, ovvero l'impiego di un linguaggio rispettoso ed inclusivo nei confronti di tutti i generi e delle diverse identità di genere¹, rappresenta uno strumento essenziale ai fini del raggiungimento della parità di genere (Diewald & Steinhauer, 2022:10).

L'imperativo della parità di diritti tra uomini e donne, del resto, è sancito dalla Legge Fondamentale, ovvero della costituzione tedesca:

Männer und Frauen sind gleichberechtigt. Der Staat fördert die tatsächliche Durchsetzung der Gleichberechtigung von Frauen und Männern und wirkt auf die Beseitigung bestehender Nachteile hin.²

(Legge Fondamentale per la Repubblica Federale di Germania, 1949)

Il *Gendern* è un processo linguistico, una strategia comunicativa, che permette di raggiungere la parità a livello linguistico (Diewald & Steinhauer, 2022:52), realizzando dunque anche nell'ambito del linguaggio e della comunicazione quanto già sancito dalla legge.

¹ Se inizialmente la lotta per la parità di genere e per l'uso di un linguaggio inclusivo riguardava solo la divisione binaria uomo-donna, col tempo la discussione si è ampliata fino ad abbracciare, ad oggi, anche identità di genere al di fuori dello schema binario (Kotthoff & Nübling, 2024:17).

² It: "Gli uomini e le donne sono equiparati nei loro diritti. Lo Stato promuove la effettiva attuazione della equiparazione di donne e uomini e agisce per l'eliminazione delle situazioni esistenti di svantaggio."

Il termine *Gendern* viene dalla parola inglese *gender* (it: genere), che in inglese indica il genere sociale, ovvero l'insieme di ruoli sociali e delle caratteristiche che stereotipicamente vengono associate ad una persona in quanto “donna” o “uomo”, e che non necessariamente corrisponde al genere biologico (*Sexus*). *ivi*, p.:10)

A questo proposito, per poter affrontare gli argomenti trattati nei prossimi capitoli è necessario dedicare una piccola parentesi a un'importante distinzione, ovvero quella tra genere sociale (*Gender*), genere biologico (*Sexus*), genere grammaticale (*Genus*) e genere semantico.

Il genere sociale (o *Gender*), come già accennato, comprende l'insieme di aspettative, ruoli di genere e stereotipi, relativamente a cosa donne e uomini pensano, a come si comportano, a quello che possono o devono fare – in sostanza, stabiliscono quello che significa essere donna o essere uomo.

Il genere biologico (*Sexus*) si riferisce invece a caratteristiche fisiche dell'individuo, che sono generalmente presenti dalla nascita. (Diewald & Steinhauer, 2022: 53).

Il genere semantico ha a che fare con i significati contenuti nelle parole, ovvero con il significato “femminile” in sostantivi come *Frau*, *Mädchen*, *Stute* (it: donna, bambina, giumenta) e il significato “maschile” in sostantivi come *Mann*, *Junge*, *Hengst* (it: uomo, bambino, stallone), a prescindere dal loro genere grammaticale (*ivi*, p. 57).

Il genere grammaticale o *Genus*, nelle lingue che lo presentano³, come il tedesco, è una caratteristica intrinseca di ogni sostantivo. In tedesco tutti i sostantivi hanno un genere grammaticale: maschile, femminile o neutro. Il *Genus* stabilisce una relazione di dipendenza grammaticale (accordo) tra gli elementi della frase (Hellinger & Bußmann, 2001: 7) e si manifesta attraverso forme linguistiche come ad esempio le desinenze (Diewald & Steinhauer, 2022: 58).

Nel caso dei *Personenbezeichnungen*, ovvero i termini che si riferiscono a persone, la categoria grammaticale e quella biologica del genere talvolta corrispondono, talvolta invece si verificano discrepanze evidenti. Un esempio è il sostantivo *das Mädchen*, che, pur essendo grammaticalmente neutro (*Genus*) in quanto derivato da un diminutivo, ha un referente femminile (*Gender + Sexus*).

³ Esistono alcune lingue prive di genere grammaticale, che permettono riferirsi a persone senza dover specificare il *Genus*. Nelle lingue a genere grammaticale, invece, come il tedesco, è necessario scegliere un *Genus* quando si fa riferimento a persone (Corbett, 1991; Gygax et al, 2019; Stahlberg et al, 2007, citato in Körner et al., 2024).

L'uso del maschile per riferirsi a gruppi misti, anche detto maschile generico, rappresenta un ulteriore esempio in cui il genere biologico dei referenti (misto) non corrisponde al genere grammaticale impiegato (esclusivamente maschile).

Il principale mezzo per esprimere un riferimento al genere biologico in tedesco, secondo il *Leibniz-Institut für Deutsche Sprache* (2024), non è il *Genus* in sé, bensì la *Movierung*, cioè la derivazione, che permette di ottenere forme femminili a partire da forme maschili (la maggior parte dei casi, come ad esempio *Student / Studentin*), o viceversa forme maschili a partire dai femminili (fenomeno più raro; un esempio è *Hexe / Hexerich*) e permette una marcatura esplicita del sesso biologico.

1.1.2 La binarietà di genere e la terza opzione

La divisione delle persone in due generi, che si riflette nel sistema linguistico, è una categorizzazione sociale e cognitiva fondamentale che affonda le sue radici nella storia. Le strutture della lingua tedesca categorizzano in maniera binaria le persone adulte nelle due categorie “uomo” e “donna”, mentre una terza categoria non è prevista dal sistema linguistico (Diewald & Steinhauer, 2022: 55-56).

Nel corso del tempo, il sistema binario che vedeva soltanto il maschile e il femminile tra i generi possibili è stato superato e progressivamente ampliato. La scienza e la medicina hanno dimostrato che a livello biologico non esiste una distinzione binaria così netta, bensì una grande varietà di forme intermedie dal punto di vista anatomico, cromosomico e ormonale. Oggi vediamo come persone intersex o con identità di genere al di fuori della binarietà convenzionale hanno ottenuto maggiori diritti e maggior rispetto (Kotthoff & Nübling, 2024:17).

Nell'Ottobre del 2017 la corte costituzionale federale tedesca ha introdotto la cosiddetta *terza opzione*, ovvero la possibilità di indicare, oltre a “*weiblich*” (it: donna) e “*männlich*” (it: uomo), anche “*divers*” (it: altro) come genere nei documenti ufficiali. L'introduzione della terza opzione ha rappresentato un grande passo in avanti per la visibilità delle identità di genere al di fuori dello schema binario uomo-donna.

Questa decisione rimane tutt'ora abbastanza controversa: la stessa Luise Pusch, nell'intervista oggetto del presente studio, la critica pesantemente, in quanto la scelta autonoma del genere si presterebbe ad abusi, consentendo l'accesso autorizzato a torto a spazi ‘riservati’ ritenuti sicuri per determinati gruppi.

Und ich denke, da sind sie dann einen Schritt zu weit gegangen, einen entscheidenden Schritt, wenn praktisch jeder, ich sage jetzt also absichtlich jeder, einfach hingehen kann

zum Standesamt und sagen kann, ich bin jetzt eine Frau. [...] Dann hat das für den Frauensport auch schon negative Konsequenzen. [...] Und was eben auch nicht geht, ist, dass Sexualverbrecher und Exhibitionisten einfach dieses Gesetz [...] missbrauchen und sich zu Frauen erklären und dann Zutritt verlangen zu geschützten Frauenräumen, zur Frauensauna. Besonders gefährlich ist es in Frauengefängnissen.⁴

(*Alles Gesagt?*, 22.11.2022, 01:38:28– 01:39:49, Pusch)⁵

Inoltre, le limitazioni della lingua tedesca (che non prevede una terza opzione dal punto di vista grammaticale) combinate ad una mancanza di terminologia consolidata per riferirsi a queste persone fanno sì che nell'interazione quotidiana col prossimo gli venga comunque assegnato un genere, che lo vogliano o meno (Diewald & Steinhauer, 2022: 62-63).

Ich würde sagen im Wesentlichen, also die Intersexuellen, die haben tatsächlich ein sprachliches Problem, weil unsere Sprache sieht eben nur zwei Geschlechter vor, und die passen weder in das eine noch in das andere. Also mein Enkelkind zum Beispiel ist non-binary und benutzt die Pronomina they und their, was im Deutschen schon wirklich kompliziert ist. Ja, im Deutschen haben wir dann noch keine richtige Lösung dafür.⁶

(*Alles Gesagt?*, 22.11.2022, 01:43:10– 01:44:29, Pusch)

Personaggi pubblici come Kim de l'Horizon cercano di smuovere qualcosa in questo senso. Nel 2022 Kim de l'Horizon ha pubblicato il libro *Blutbuch* aggiudicandosi i premi per il miglior libro tedesco (*Deutscher Buchpreis*) e svizzero (*Schweizer Buchpreis*). La figura principale all'interno del libro è una persona non binaria che utilizza un linguaggio proprio, praticamente privo di riferimenti al genere.

⁴ It: "E penso che in questo caso abbiano fatto un passo di troppo, un passo determinante, se praticamente ognuno, e non a caso dico ognuno, può semplicemente recarsi all'anagrafe e dichiarare di essere una donna. [...] Questo ha conseguenze negative anche per lo sport femminile. [...] E poi non è possibile che i criminali sessuali e gli esibizionisti abusino di questa legge [...] dichiarandosi donne e chiedendo poi di accedere a spazi riservati alle donne, come la sauna femminile. La questione è particolarmente pericolosa nelle carceri femminili."

⁵ Le citazioni che riportano come fonte "*Alles Gesagt?*, 22.11.2022" sono tratte dall'intervista oggetto di questo studio (cf. cap. 2) e riportano il nome del podcast, la data dell'episodio, il minutaggio corrispondente allo spezzone riportato in citazione e la persona a cui la citazione appartiene.

⁶ It: "Direi sostanzialmente che gli intersessuali hanno effettivamente un problema linguistico, perché la nostra lingua prevede solo due generi e loro non rientrano né nell'uno né nell'altro. Ad esempio, mi* nipote è non-binary e usa i pronomi *they* e *their*, il che in tedesco è davvero complicato. Sì, in tedesco non abbiamo ancora una soluzione adeguata per questo".

1.2 Il maschile generico e il *Male As Norm*

Il maschile generico viene utilizzato in diverse lingue a genere naturale o a genere grammaticale per riferirsi a persone o gruppi di persone il cui genere è misto o irrilevante (Hellinger & Bußmann, 2001:9). A partire dalle critiche mosse dalla linguistica femminista negli anni Settanta, diversi studi hanno riscontrato che il genere grammaticale utilizzato influisce sulle rappresentazioni mentali di chi lo recepisce, e che il maschile generico in particolare produce un *bias* maschile. (Klein, 1988; Gygax et al., 2008; Stahlberg et al., 2007).

Per *bias* maschile si intende che le immagini mentali prodotte dall'ascoltare un termine al maschile generico (ad esempio "gli studenti" invece che "le studentesse e gli studenti") tendono ad essere maschili, piuttosto che femminili o neutre. In sostanza, le persone tendono a immaginare maggiormente uomini piuttosto che donne. L'impiego di forme che includono e rappresentano anche le donne, al contrario, porta a una maggiore visibilità delle stesse e ad una maggiore quota di donne nelle rappresentazioni mentali di chi ascolta (Stahlberg et al., 2001; Sczesny et al., 2016; Körner et al., 2024; Glim et al., 2025).

Il maschile generico è, a livello linguistico, un'espressione del *Male As Norm* (MAN), l'ideologia alla base del patriarcato, che vede l'uomo come elemento centrale e privilegiato, e la donna come ausiliare, secondaria (Diewald & Steinhauer, 2022: 17).

1918 bekamen die Frauen in Deutschland das Wahlrecht. Seitdem gibt es neben den »Wählern« auch »**weibliche** Wähler«, neben den »Parlamentariern, Politikern, Staatssekretären und Ministern« auch »**weibliche** Parlamentarier, Politiker, Staatssekretäre« [...]. Wir sehen: Das Neue, Ungewöhnliche, nicht der Norm Entsprechende wird jeweils sprachlich gekennzeichnet.⁷

(Pusch, 1984:101, grassetto mio.)

La critica linguistica femminista a partire dagli anni Settanta denuncia il *Male As Norm* (cfr. Lakoff, 1973), che contribuisce ad una visione androcentrica del mondo. Secondo la logica del MAN, le donne dovrebbero semplicemente "sentirsi incluse" anche quando non vengono esplicitamente menzionate. Ma siccome la forma maschile generica coincide in tutto e per tutto con la forma maschile specifica, usata per riferirsi a sole persone di sesso maschile, diventa

⁷ It: "Nel 1918 le donne in Germania hanno ottenuto il diritto di voto. Da allora, oltre agli «elettori» esistono anche «gli elettori **donna**», oltre ai «parlamentari, politici, sottosegretari e ministri» esistono anche «parlamentari, politici, sottosegretari **donne**» [...]. Vediamo come ciò che è nuovo, insolito, non conforme alla norma viene sempre contrassegnato linguisticamente."

virtualmente impossibile distinguere quando le donne possono o devono sentirsi incluse o escluse. Questo riduce l'inclusione o esclusione delle donne ad un'*implicatura conversazionale*, che può essere in qualsiasi momento negata, rimossa o eliminata (Diewald & Steinhauer, 2022: 25-26).

Sono diverse anche le voci che si schierano a favore del cosiddetto maschile generico, sostenendo che abbia un valore neutrale e negandone gli effetti discriminatori.

In realtà, già dal Rinascimento il maschile plurale era utilizzato in senso non neutrale ma specificatamente riferito al genere maschile; l'idea che abbia valore neutrale è relativamente nuova e risale alla seconda metà del ventesimo secolo. Se in passato il maschile veniva usato miratamente, ad esempio per escludere le donne da determinate cariche, a partire dal ventesimo secolo alla stessa forma viene conferito un valore "neutrale" per evitare di menzionare le donne, con annesse le oggi note conseguenze sulla loro visibilità e sulle rappresentazioni mentali (ivi, p.23-24).

Pusch evidenzia come il maschile generico suoni particolarmente inadatto e "grottesco" in contesti prettamente femminili, come quello della gravidanza e dell'allattamento: ne sono chiari esempi frasi come „Wenn man sein Kind stillt..." (it: "Quando **uno** allatta il suo bambino...") o „Wie kann **man** seine Schwangerschaft frühzeitig selbst feststellen?" (it: "Come fa **uno** a riconoscere presto la propria gravidanza?") (Pusch, 1984:86, grassetti miei).

L'abitudine di ricorrere al maschile anche laddove esistono equivalenti femminili porta anche a situazioni paradossali. Luise Pusch mostra il paradosso del maschile sovraesteso con un esempio lampante: il foglietto illustrativo degli assorbenti interni della O.B. recitava „Die Menstruation ist bei **jedem** ein bißchen anders." (it: "Le mestruazioni sono leggermente diverse per ognuno"). Sotto richiesta della linguista stessa, la frase è stata modificata in „Die Menstruation ist bei **jeder Frau** ein bißchen anders." (it: "Le mestruazioni sono leggermente diverse per ogni **donna**") (ivi, pp.149-150; grassetti miei).

1.3 La linguistica femminista

Le radici della linguistica femminista affondano negli Stati Uniti. Nel 1970, durante il congresso dell'*American Dialect Society*, Mary Ritchie Key tiene il primo seminario sul tema Lingua e Genere: *Linguistic Behavior of Male and Female* (Kotthoff & Nübling, 2024:20). Il seminario, insieme alla successiva pubblicazione del suo libro *Male/Female Language* (1975) apre la strada al movimento. In quegli anni viene pubblicato anche *Language and Woman's*

place (1973) di Robin Lakoff, che mette in luce la marginalizzazione delle donne a livello linguistico e l'annullamento della donna per via del maschile generico.

La linguistica femminista in quegli anni evidenzia le forme di subordinazione del cosiddetto linguaggio femminile rispetto al linguaggio maschile. Il linguaggio femminile si contraddistingue per la presenza di molti segnali discorsivi (ad es. “credo, mi pare, non so” ecc.), continue richieste di conferma (come “non è vero?”), maggiore cortesia, emotività esasperata, minore competenza lessicale (Bazzanella, 2010). Il linguaggio femminile, caratterizzato dalla sottomissione, dal contenimento e dall'incertezza, sottolinea la subordinarietà delle donne a livello sociale (Kotthoff & Nübling, 2024:20).

Già nel 1972 anche Casey Miller e Kate Swift avevano denunciato la tendenza all'uso del maschile per i sostantivi legati a professioni e l'uso del pronome maschile *he/him* a uso generico, e nel loro *Handbook for Nonsexist Writing* (1980) proponevano alcune soluzioni per evitare formulazioni sessiste che oggi sono ampiamente usate in inglese, come il pronome ‘generico’ *they* per la terza persona singolare.

Se i primi studi femministi erano volti ad evidenziare le forme di subordinazione del cosiddetto ‘linguaggio femminile’, in seguito si cominciò ad analizzarlo in prospettiva opposta, cioè sottolineandone i punti di forza. Il linguaggio femminile era caratterizzato infatti da “maggiore cooperatività, coinvolgimento, disponibilità alla negoziazione, capacità di ascolto e di ripresa di quanto detto dall'interlocutore” (Bazzanella, 2010). Successivamente, la netta distinzione tra linguaggio maschile e femminile venne superata, constatando che le strategie linguistico-comunicative tipiche del linguaggio femminile sono in realtà tipiche di qualsiasi soggetto che si trovi a ricoprire una posizione subordinata all'interno dell'interazione, indipendentemente dal genere (ibid.).

Diversi studi hanno dimostrato l'esistenza di una disparità conversazionale tra donne e uomini: le donne tendono ad avere uno status sociale inferiore, e a ciò che fanno viene data meno considerazione rispetto agli uomini. È stato riscontrato, ad esempio, che essere intervistati da una donna è sinonimo di minore prestigio per gli intervistati (Trömel-Plötz, 1984:354, citato in Erfurt, 1988); che l'autorità e la credibilità dei telegiornali viene messa in dubbio, se le notizie vengono lette da donne; che gli interventi delle donne nei dibattiti televisivi vengono interrotti più spesso dal moderatore; che alle oratrici donne viene concesso meno tempo di parola rispetto agli uomini (Kramarae 1984, citato in Erfurt, 1988).

In Germania, il filone della critica linguistica femminista viene ripreso da Luise Pusch e Senta Trömel-Plötz, che insieme a Marlis Hellinger sono considerate co-fondatrici della linguistica femminista tedesca (Olderdissen, 2019).

Interviewer: Also unsere Gästin von heute [...] ist die Begründerin der feministischen Linguistik in Deutschland.

Luise Pusch: Die – ist wirklich wichtig zu sagen – Mitbegründerin. Da sind noch zwei andere, die wesentlich mitgewirkt haben.

Interviewer: Senta Trömel-Plötz und Marlis Hellinger?

Luise Pusch: Genau.⁸

(*Alles Gesagt?*, 22.11.2022: 00:08:42 – 00:09:18, Intervistatore & Pusch)

Nel 1978, Senta Trömel-Plötz è la prima in ambito tedesco a proporre la discriminazione linguistico-strutturale delle donne come legittimo oggetto di ricerca linguistica: in *Linguistik und Frauensprache*, la linguista critica l'uso del maschile generico e propone soluzioni alternative, come lo sdoppiamento, che consiste nel menzionare sempre donne e uomini quando si parla di gruppi misti invece di usare il maschile generico, garantendo visibilità alle donne e creando così rappresentazioni mentali più eque.

L'anno dopo Hartwig Kalverkämper pubblica *Die Frauen und die Sprache*, una critica in cui attacca l'analisi di Trömel-Plötz definendola “*unwissenschaftlich*” (it: “ascientifica”) (Kalverkämper, 1979). Luise Pusch contrattacca con la celebre risposta *Der Mensch ist ein Gewohnheitstier, doch weiter kommt man ohne ihr: Eine Antwort auf Kalverkämpers Kritik an Trömel-Plötz* (1979), in cui argomenta la necessità di una critica linguistica in chiave femminista che si confronti col tema delle discriminazioni linguistiche strutturali che rendono le donne invisibili.

Marlis Hellinger in *Sprachwandel und feministische Sprachpolitik. Internationale Perspektiven* (1985:3-4) offre una panoramica plurilinguistica, che evidenzia come diverse lingue posseggano meccanismi analoghi di oscurazione del genere femminile. Hellinger identifica le seguenti costanti:

⁸ It: “Intervistatore: Dunque la nostra ospite di oggi [...] è la fondatrice della linguistica femminista in Germania. Luise Pusch: La – è importante precisarlo – co-fondatrice. Ci sono altre due donne che hanno contribuito significativamente.

Intervistatore: Senta Trömel-Plötz e Marlis Hellinger?
Luise Pusch: Esatto.”

- 1- Gli sforzi e le conquiste delle donne non vengono rappresentati a livello sociale per via dell'uso del maschile generico;
- 2- Esiste un'asimmetria semantica (denotativa e connotativa) tra espressioni maschili e femminili. Le forme femminili tendono ad avere connotazione dispregiativa o diminutiva o comunque a godere di un rango inferiore rispetto a quelle al maschile;
- 3- Il riferimento ad una donna al maschile viene interpretato come una valorizzazione mentre riferirsi a un uomo con una forma al femminile ha valore dispregiativo.

Negli anni si sono susseguiti importanti saggi e raccolte di diverse linguiste: Bußmann, Schoenthal, Diewald e Nübling per citarne alcune. Samel (2000) ha realizzato la prima introduzione alla linguistica femminista (*Einführung in die feministische Sprachwissenschaft*), seguita poi da quella realizzata da Klann-Delius (2005). Ayaß (2008) con *Kommunikation und Geschlecht* fornisce un'introduzione alla sociologia della comunicazione. *Das Deutsche als Mönnersprache* di Luise Pusch (che approfondiremo nella prossima sezione) è considerato una pietra miliare della linguistica femminista tedesca (Kotthoff & Nübling, 2024:21).

1.3.1 Luise Pusch – *Das Deutsche als Mönnersprache*

Nel 1984 Luise Pusch pubblica la raccolta di saggi *Das Deutsche als Mönnersprache*. L'autrice illustra le idee ed obiettivi del *Neue Frauenbewegung* (it: nuovo movimento femminista) e descrive alcune proposte volte a “femminilizzare” la lingua: secondo l'autrice ci sarebbe bisogno di una vera e propria *Rettungsaktion* (azione di salvataggio) per contrastare l'annullamento linguistico della donna imposto dalla società maschilista (Pusch, 1984: 8-11).

Il *Neue Frauenbewegung* mette in luce in chiave critica la predominanza del maschile a livello linguistico e propone soluzioni per superarla. Le linguiste femministe si oppongono alla regola dettata dalla cultura patriarcale “*weiblich gleich zweitrangig*” (it: “donna = secondaria”). Conseguenza di questa regola patriarcale è, ad esempio, che la femminilizzazione equivale a un declassamento, mentre le donne possono essere “elevate” attraverso la mascolinizzazione. Per questo motivo, frasi come “Paul Celan era una Ingeborg Bachmann/Nelly Sachs uomo” non possono essere considerate come lodi, mentre è molto comune la formulazione opposta, del tipo: “Erica Jong è un Henry Miller donna” (ivi, pp. 76-78).

Pusch argomenta che l'unico modo per rompere questo meccanismo è *spatriarcalizzare* (ivi, p. 86) la lingua è la valorizzazione del femminile attraverso il suo utilizzo coerente, consapevole e intensificato: il femminile generico, che sovverte l'ordine patriarcale, opponendosi esplicitamente al maschile generico (ivi, pp. 76-78).

Il femminile generico nasce come provocazione in seno al nuovo movimento femminista. Così come il suo corrispettivo maschile, il femminile generico viene utilizzato per rivolgersi a un pubblico di genere misto. La proposta è, in sostanza, di rovesciare l'ordine patriarcale delle cose e parlare al femminile, anche quando si parla di uomini, per “rimediare” a secoli di linguaggio androcentrico.

Was ich vorgeschlagen habe, war eben einfach die Sache umzudrehen. [...] Also einfach im Femininum reden, die Männer sind immer alle herzlich mitgemeint.⁹

(*Alles Gesagt*, 22.11.2022, 01:41:15 - 01:43:09, Pusch)

Ad oggi, il femminile generico non è riuscito ad affermarsi con un vero valore “generico” nel senso di “neutrale”, in quanto questa forma provoca un *bias femminile*, che, per quanto contrasti il *bias maschile* prodotto dal maschile generico, non è in grado da solo di fornire rappresentazioni mentali equilibrate (Körner et al., 2024).

Il processo di “femminilizzazione” del linguaggio interessa anche altri aspetti del sistema linguistico. Pusch descrive l'utilizzo dei (neo)pronomi al posto dell'usuale pronome indefinito *man*, come *frau* e *eine*¹⁰ per riferirsi a persone di genere femminile, ma anche *mann* per riferirsi a persone di genere maschile, e *mensch*, utilizzato in casi in cui il genere è irrilevante (ivi, pp. 86-93). Viene descritta anche la possibilità di ripresa del pronome *wer* al femminile („wer, die/ihr...“) invece che al maschile („wer, der/sein...“) (Pusch, 1984: 89-91).

Tra le strategie proposte dal movimento rientra anche lo sdoppiamento, e viene tematizzata la produttività del suffisso *-in* e del morfema *-frau* (ad es. *die Geschäftsfrau*, *die Kauffrau*) nella creazione di alternative femminili ad alcune professioni che in passato esistevano solo al maschile perché esercitate esclusivamente o prevalentemente da uomini.

Il nuovo movimento femminista ha elaborato anche soluzioni più radicali, spesso come forma di provocazione o anche di gioco. Le regole di congruenza femministe radicali prevedono la trasformazione di tutti i sostantivi riferiti a persone nella forma femminile, anche all'interno di una parola: così *freundlich* (it: amichevole) diventa *freundinlich*, *Freundschaft* (it: amicizia)

⁹ It: “Quello che ho proposto era semplicemente di ribaltare la situazione. [...] Quindi semplicemente parlare al femminile, gli uomini sono sempre cordialmente inclusi.”

¹⁰ Il pronome indefinito *eine* (“una”) è la controparte femminile di *einer* (“uno”). Le forme del pronome *einer* all'accusativo (*einen*) e al dativo (*einem*) fungono anche da forme suppletive per il pronome *man*, che esiste in questa forma solo al nominativo. Le forme *man/frau* e *einer/eine* sono forme concorrenti al nominativo. Per approfondimenti sul pronome indefinito *man* e sulle sue alternative si rimanda a Zifonun (2000) e Storzjohann (2004).

diventa *Freundinnenschaft* e *sich anfreunden* (it: fare amicizia) diventa *sich anfreundinnen* (ivi, p. 105).

Tra le proposte più “estreme” del movimento troviamo anche l’abolizione totale della *Movierung* effettuata tramite il suffisso *-in*, in quanto la creazione del femminile a partire dalla forma maschile “preserva nel sistema linguistico la millenaria dipendenza delle donne dagli uomini, che deve essere finalmente superata anche nella lingua.

Statt die Studentin/Lehrerin/Ingenieurin einfach immer die Student/Lehrer/Ingenieur, etc.!¹¹

(ivi, p. 61)

1.4 Lingua, pensiero e marginalizzazione

Alla base della critica linguistica femminista c’è il legame tra linguaggio, pensiero, e realtà. Nel 1989 Gisela Schoenthal descrive la lingua come al contempo specchio, ostacolo e strumento per e nel rappresentare, comprendere e plasmare la realtà:

Sprache einerseits als Spiegel, als Ausdruck historisch gewachsenen Denkens, Sprache andererseits als Hindernis, eine sich wandelnde oder schon gewandelte Wirklichkeit wahrzunehmen, Sprache aber auch als Hilfsmittel, an dieser Wandlung mitzuwirken.¹²

(Schoenthal, 1989:300)

La lingua non è quindi solo un riflesso della realtà ma definisce a sua volta la realtà stessa e rappresenta uno strumento importante per la lotta alle disuguaglianze politiche e di genere.

Kotthoff e Nübling spiegano come le rappresentazioni del genere (il cosiddetto *doing gender*) non hanno a che fare solo con come ci vestiamo, quale professione svolgiamo e così via, ma anche e soprattutto con la lingua, il modo in cui interagiamo e persino con le strutture grammaticali della lingua stessa. Se da un lato la lingua crea il sedimento, la base su cui si sviluppa la nostra percezione (e di conseguenza la nostra realtà), dall’altro nel parlare, in ogni atto linguistico, attiviamo queste stesse categorizzazioni mentali e le rinforziamo:

¹¹ It: “Invece di la studentessa/professoressa/ingegnera diciamo semplicemente la studente/professore/ingegnere, eccetera!”

¹² It: “Da un lato la lingua come specchio, come espressione di un pensiero sviluppatosi nel corso della storia; dall'altro lato la lingua come ostacolo alla percezione di una realtà in trasformazione o già trasformata; ma anche la lingua come strumento per contribuire a questa trasformazione.”

So sind Ausdrücke wie Köchin, Arzt, sie, er nicht nur bloße Referenzformen, sondern gleichzeitig (je nach Sichtweise auch ausschließlich) sog. Appellationen mit wirklichkeitskonstituierender Funktion (Hornscheidt 1998, 2006).¹³

(Kotthoff & Nübling, 2024:21-22)

I sistemi linguistici sono frutto della sedimentazione, nel corso dei secoli, del modo in cui i due generi parlavano l'uno con l'altro o l'uno dell'altro. Secondo Kotthoff e Nübling, il genere è talmente presente nella lingua tedesca, che è praticamente impossibile evitarlo (ivi, p. 21). Il sistema linguistico, con le sue strutture sedimentate, conserva, perpetua e riproduce i modelli di genere del passato (ivi, p. 16).

Secondo Haß-Zumkehr (2003:162), nella lingua ritroviamo non tanto un'immagine o una rappresentazione, quanto più un'impronta della società che continua a influenzare le nostre percezioni, fino a quando qualcosa non cambia a livello della nostra mentalità o a livello della società. Il cambiamento sociale e quello linguistico possono scaturire l'uno dall'altro e influenzarsi vicendevolmente. Separare i due è spesso impossibile.

Per quanto la lingua costituisca la base delle nostre percezioni, però, essa non le determina, altrimenti il cambiamento linguistico (che avviene costantemente) sarebbe impensabile (Kotthoff & Nübling, 2024:21).

Luise Pusch considera il cambiamento linguistico non solo possibile, ma assolutamente necessario. Secondo Pusch, l'uomo ha creato e organizzato il mondo a sua immagine e somiglianza. La presenza di un altro sesso disturbava quest'armonia. Per ristabilire l'ordine, si è fatto ricorso alla grammatica: così, per parlare di novantanove maestre ed un maestro si dice "cento maestri". Alle donne (in questo caso, alle maestre) rimangono due opzioni: o considerarsi sottintese nel genere maschile, o non esistere affatto. La donna di oggi non si accontenta di essere "inclusa" nelle espressioni al maschile o di non essere menzionata affatto, e comincia invece a costruire un mondo a propria immagine e somiglianza, creando una nuova armonia, con l'obiettivo di creare un mondo che sia in armonia con entrambi i generi: non più un mondo degli uomini, ma di tutti gli esseri umani (ivi, pp. 106-107. Sottolineatura mia.).

¹³ It: "Espressioni come cuoca, medico, lei, lui non hanno quindi soltanto una funzione referenziale: sono allo stesso tempo (a seconda del punto di vista anche esclusivamente) cosiddette appellazioni con funzione costitutiva della realtà (Hornscheidt 1998, 2006)."

La linguistica femminista propone due strategie per raggiungere questo obiettivo: le strategie di visibilità, che consistono nel menzionare esplicitamente le donne invece di ricorrere al maschile generico, e la neutralizzazione, ovvero l'eliminazione di ogni riferimento al genere.

1.5 Reazioni e sviluppi: le prime linee guida

Fin dai primi passi mossi negli anni Settanta, le proposte femministe si sono scontrate con reazioni di resistenza e rifiuto. Il fenomeno veniva inizialmente visto come qualcosa di folle e ridicolo, seppur fondamentalmente innocuo:

die übliche Reaktion auf Neues und Fremdes, das uns "letztlich nichts angeht" [...].
Sollten "die" doch ruhig frau statt man sagen und sich damit lächerlich machen.¹⁴

(Pusch, 1984:83)

La resistenza alle innovazioni proposte, secondo Hellinger (1985:255-257), si manifesta sotto forma di quattro strategie:

1. Le linee guida e i regolamenti già esistenti in materia di linguaggio di genere vengono deliberatamente ignorati o applicati solo in parte;
2. Il cambiamento linguistico viene considerato inutile allo scopo di portare al cambiamento sociale, negando il rapporto tra linguaggio e società;
3. Alcuni fenomeni singoli vengono estrapolati dal loro contesto per essere ridicolizzati (ad esempio, alcune proposte femministe "estreme" vengono utilizzate per screditare l'intero movimento);
4. L'utilizzo di forme inclusive viene scoraggiato in quanto infrangerebbe non ulteriormente precisati principi di correttezza, economia o persino estetica.

Gli oppositori sostengono ad esempio l'idea che non avrebbe alcun senso lottare per l'uguaglianza a livello linguistico se non si è raggiunta l'uguaglianza a livello politico, economico, sociale:

Ob es etwas bringt, wenn die Sprache geändert wird und die Verhältnisse bleiben die gleichen? Mir ist es lieber, die Sprache bleibt in ihren [...] eingefahrenen Gleisen seit dem 11. Jahrhundert und die Verhältnisse ändern sich stattdessen. Wenn alle die Energien, die darauf verschwendet werden und wurden, daß »man« durch »frau« ersetzt

¹⁴ It: "La reazione abituale nei confronti di ciò che è nuovo e sconosciuto, e cioè che "in fin dei conti non ci riguarda" [...]. Che "loro" continuino pure a dire *frau* invece di *man* e a rendersi ridicole."

wird, dazu verwendet worden wären, unsere Gesellschaft für Kinder, Frauen und Männer erträglicher zu machen, dann wären wir vielleicht schon einen kleinen Schritt weiter.¹⁵

(Kolvenbach, 2019)

L'idea alla base di questa critica è che una volta cambiata la realtà, la lingua cambierebbe da sé. Pusch (1984:83) contraddice questa argomentazione affermando che, innanzitutto, la lingua è oggi l'ambito centrale della realtà, e che del resto nessuna linguista femminista ha mai affermato di voler cambiare solo la lingua e lasciare che tutto il resto nelle mani del patriarcato. Inoltre, se da un lato gli stereotipi e i ruoli di genere esistono anche indipendentemente dalla lingua, questi vengono rafforzati se utilizziamo un linguaggio che li conferma e li sottolinea invece di combatterli (Diewald & Steinhauer, 2022:103).

Nella sua tesi *Linguaggio di genere e interpretazione: un connubio possibile? Un contributo dall'analisi di linee guida istituzionali in italiano, inglese e tedesco* (2024) Sofia Sorbelli ripercorre il susseguirsi di linee guida, risoluzioni e proposte di linguaggio di genere in italiano, inglese e tedesco dagli anni Ottanta ai giorni nostri. Nell'ambito della presente tesi, che si focalizza sul linguaggio di genere in tedesco, riportiamo brevemente i punti salienti di questa carrellata storica.

Le prime linee guida per un uso non sessista del tedesco furono pubblicate nella Repubblica Federale Tedesca nel 1980. *Richtlinien zur Vermeidung sexistischen Sprachgebrauchs* di Ingrid Guentherodt, Marlis Hellinger, Luise F. Pusch e Senta Trömel-Plötz è una delle prime proposte in questo senso: presenta diversi esempi di linguaggio sessista a cui corrispondono soluzioni alternative non sessiste.

Nel 1987 Ruth Wodak, Gert Feistritz, Sylvia Moosmüller e Ursula Doleschal presentano il progetto *Mann und Frau im öffentlichen Bereich. Linguistische Richtlinien zur geschlechtsneutralen Sprachverwendung*, che propone diverse strategie per garantire visibilità e simmetria, tra cui la neutralizzazione tramite astrazione (trasformando ad esempio *der Leiter*, il direttore, in *die Leitung*, la direzione), la corretta formazione dei femminili per le professioni e lo sdoppiamento. Nel 1990, viene pubblicato il rapporto *Maskuline und feminine*

¹⁵ It: "Che senso ha cambiare la lingua se le condizioni rimangono invariate? Preferisco che la lingua rimanga nei suoi binari [...] consolidati dall'XI secolo e che invece cambino le condizioni. Se tutte le energie che sono state e vengono sprecate per sostituire «man» con «frau» fossero state utilizzate per rendere la nostra società più vivibile per bambini, donne e uomini, forse avremmo già fatto un piccolo passo avanti."

Personenbezeichnungen in der Rechtssprache, con focus in particolare sul linguaggio di genere in ambito giuridico (Parlamento tedesco, 1990).

A livello internazionale, nel 1993 è stata emanata dall'UNESCO una risoluzione per sensibilizzare la Direzione Generale sull'uso di un linguaggio non sessista, che rendesse le donne più visibili. Sulla base di questa risoluzione, Marlis Hellinger e Christine Bierbach hanno poi redatto le linee guida per un uso non sessista della lingua (1993) per conto della commissione tedesca dell'UNESCO di Bonn. Tra le proposte troviamo diversi esempi interessanti, come l'uso simmetrico di cognome e nome o solo cognome per donne e uomini (ad esempio, Sartre und de Beauvoir e non Sartre und Simone de Beauvoir) (Hellinger & Bierbach, 1993). Negli anni si sono seguite altre pubblicazioni con il medesimo obiettivo anche in altri paesi di lingua tedesca. Ad esempio, in Svizzera vennero introdotte delle norme ufficiali pubblicate dalla Cancelleria federale svizzera nel 1996.

Col passare degli anni, il numero di manuali e linee guida è cresciuto esponenzialmente. Tra le pubblicazioni degli ultimi anni possiamo citare ad esempio la linea guida pubblicata nel 2020 dalla Società per la lingua tedesca (Gesellschaft für die deutsche Sprache, 2020), quella pubblicata dalla Cancelleria Svizzera nel 2023 (Cancelleria federale svizzera, 2023) e, in ambito austriaco, la linea guida *Geschlechtergerechte Sprache*, pubblicata nel 2024 dalla Cancelleria federale austriaca.

Sono numerose anche le pubblicazioni in ambito accademico e universitario, dove troviamo le strategie più all'avanguardia rispetto al linguaggio inclusivo, che tendono maggiormente verso l'inclusione delle persone non binarie. Tra queste, la linea guida *Geschlechtergerechte Sprache* della sezione *Diversity* dell'Università Heinrich Heine di Düsseldorf propone tra le varie strategie anche l'uso dell'asterisco, del trattino basso o dei due punti all'interno delle parole (2024:51), mentre la guida dell'Università di Colonia *überzeugENDERe Sprache* (Universität zu Köln, 2021:20) ammette anche la possibilità di utilizzare il pronome indefinito *mensch* al posto di *man*.

Per quanto sia evidente che il dibattito intorno alla questione del genere sia molto acceso e vivace, la proliferazione di linee guida differenti e disomogenee ha portato ad una sensazione di scoraggiamento e di disorientamento per le persone comuni che tentano di cimentarsi con il *Gendern*.

Zifonun teme che si arrivi ad una sorta di “slalom” di soluzioni impiegate per evitarne altre. L'autrice spiega che la lingua va accettata (e amata) come sistema imperfetto e per certi versi

ingiusto, applicando riforme linguistiche moderate e ponderate rispetto al genere (Zifonun 2018: 48).

Karin Wetschanow (2017) parla a questo proposito di un rapporto di tensione (*Spannungsverhältnis*) tra linguaggio inclusivo e linguaggio “corretto”. Questo schema vede schierati, da un lato, manuali e linee guida che si concentrano sulla „correttezza“, che generalmente evitano qualsiasi tipo di intervento sul sistema linguistico, e consigliano ad esempio lo sdoppiamento e i termini astratti o collettivi; dall’altro, chi invece punta maggiormente all’inclusività tende a proporre interventi di modifica – talvolta anche massicci – all’interno del sistema linguistico, come ad esempio l’uso di (neo)pronomi o dei tanto discussi simboli grafici¹⁶. Solitamente chi propone soluzioni più “radicali” tende ad essere mosso da obiettivi o ideologie politico-sociali (Diewald & Steinhauer, 2022:40-41).

Nell’applicare il linguaggio di genere ai bisogni della vita quotidiana non è possibile affidarsi ciecamente ad una sola regola sempre valida: è necessario soppesare le proprie scelte linguistiche consapevolmente, in base a contesti e circostanze, in modo da esprimersi in maniera al contempo rispettosa dei generi, corretta e comprensibile (ivi, p. 15). Per un’attuazione efficace, ovvero accettabile da tutti, del linguaggio di genere, è necessario valutare attentamente i limiti e soppesare i vantaggi e gli svantaggi derivanti dall’uso di determinate forme e strategie (ivi, pp. 32-33).

1.6 Il linguaggio di genere in tedesco oggi

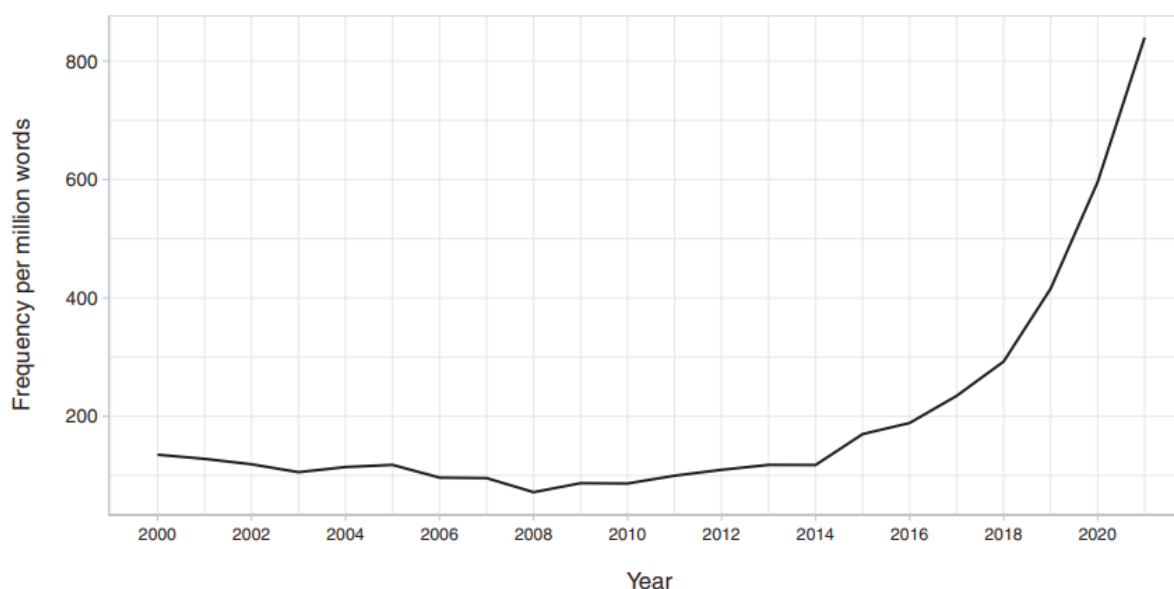
Oggi la discussione intorno alla questione del linguaggio inclusivo si è ampliata attraverso l’entrata in gioco di nuove soluzioni, nuove teorie, nuove proposte, e il conseguente moltiplicarsi di manuali e linee guida.

Negli ultimi anni in Germania si è registrato un aumento nell’utilizzo del linguaggio di genere, in particolare nel linguaggio utilizzato da media e testate giornalistiche (Waldendorf, 2024). Se da un lato le forme che includono persone non binarie tendono ad essere meno popolari e vengono utilizzate prevalentemente nelle testate di sinistra, altre forme di linguaggio inclusivo si sono diffuse rapidamente anche in testate di destra, che utilizzano il linguaggio di genere in misura minore. Tra queste, soprattutto lo sdoppiamento ‘classico’, che aderisce alla visione

¹⁶ I simboli grafici, come l’asterisco (*Gendersternchen*), il trattino basso (*Gender Gap*) e i due punti (*Doppelpunkt*) vengono criticati in quanto non conformi alle regole ortografiche vigenti.

binaria del genere ed è conforme alle regole di ortografia, ha preso piede in diverse testate a prescindere dall'orientamento politico. (ivi, pp. 367-369).

Figura 1 – Frequenza di linguaggio inclusivo di genere dal 2000 al 2021 nel corpus DeReKo (da Waldendorf, 2024: 363)



La figura 1 mostra i dati relativi alla frequenza del linguaggio inclusivo di genere rilevata all'interno del corpus DeReKo (Deutsches Referenzkorpus) tra il 2000 e il 2021 in uno studio condotto da Anica Waldendorf nel 2024. I fenomeni presi in considerazione in questo caso sono dieci tipi differenti di sdoppiamento (ivi, p. 362).

La graduale diffusione del linguaggio di genere è dovuta a diversi fattori storici e sociali. Oltre all'enorme impatto della critica linguistica femminista, anche l'ingresso del partito dei Verdi nel Bundestag nel 1983 è stato determinante. Da quel momento, il linguaggio inclusivo di genere si è infatti rapidamente diffuso anche all'interno del parlamento, dove fino ad allora era praticamente inesistente (Stecker et al., 2021:9).¹⁷

Ad oggi, il partito dei Verdi rimane in prima linea in questo senso, insieme ai partiti di sinistra e ai socialdemocratici (SPD) che mostrano una maggiore propensione per il linguaggio inclusivo rispetto al partito liberaldemocratico (FDP) e ai conservatori (CDU/CSU), mentre il partito populista di destra AfD si distingue per un livello di linguaggio inclusivo

¹⁷ Lo studio analizza l'evoluzione del linguaggio di genere nel Bundestag tedesco attraverso il corpus *GermaParl*, che contiene dibattiti parlamentari tedeschi dal 1949 a oggi (oltre 266 milioni di token). È stato costruito un dizionario delle forme maschili e femminili dei 1600 *Personenbezeichnungen* più frequenti (circa 2,5 milioni di occorrenze totali).

eccezionalmente basso. È stato inoltre osservato che l'innovazione linguistica sotto questo punto di vista è trainata dalle parlamentari donne, che impiegano il linguaggio di genere in misura maggiore rispetto ai colleghi uomini, a prescindere dall'affiliazione politica (ibid.).

In seguito all'introduzione della legge § 611 b BGB, che prevede che tutte le offerte di lavoro siano rivolte sempre esplicitamente sia a uomini che a donne, la pratica dello sdoppiamento negli annunci di lavoro si è diffusa rapidamente in Germania (Pusch, 1984:93-97)¹⁸.

Nonostante la sua rapida diffusione nei media e nella politica, il linguaggio di genere rimane un tema controverso. Tra le motivazioni che portano le persone a rifiutare il linguaggio di genere, c'è ad esempio l'idea che rendano i testi macchinosi, che la lingua diventi complicata e scoraggiante per chi sta cercando di imparare il tedesco, come anche il fatto che i testi diventino "illeggibili e più lunghi". (Schneider, 2020; Rock, 2021; Web editorial staff del LpB BW, 2023; citati in Müller-Spitzer et al., 2024). Contrariamente a quanto spesso argomentato, però, diversi studi hanno dimostrato che il linguaggio di genere non riduce la comprensibilità dei testi (Blake & Klimmt 2010; Braun et al. 2007; Friedrich & Heise 2019; Friedrich et al., 2022). Un'altra critica comune riguarda lo sforzo linguistico necessario per creare questi testi, che poi diventano "lunghi, monotoni e ossessionati dal genere" (Eisenberg, 2022, citato in Müller-Spitzer et al., 2024).

In realtà, secondo uno studio condotto da Müller-Spitzer et al. (2024), nel complesso le modifiche da apportare per rendere i testi rispettosi dal punto di vista del genere riguarderebbero meno dell'1% delle parole nei testi della stampa tedesca¹⁹ e riguarderebbero principalmente i sostantivi, in particolare quelli che si riferiscono a persone al maschile generico. I sostantivi riferiti a persone o *Personenbezeichnungen* sono infatti la classe linguistica più frequente, nonché maggiormente interessata dal linguaggio di genere e maggiormente rilevante nello studio della rappresentazione linguistica delle persone (ivi, p. 17).

Esistono inoltre diverse strategie che permettono di trasformare i sostantivi rendendoli inclusivi senza la necessità di adoperare i simboli grafici, come lo sdoppiamento e il ricorso ai sostantivi

¹⁸ La norma stabiliva che „Der Arbeitgeber (sic) soll einen Arbeitsplatz weder öffentlich noch innerhalb eines Betriebes nur für Männer oder nur für Frauen ausschreiben.“ (it: “Il datore di lavoro (sic) non deve pubblicare offerte di lavoro solo per uomini o solo per donne né pubblicamente né all'interno di un'azienda”) (Pusch, 1984:95). La legge non è più in vigore in quanto nel 2023 è stata integrata nell' *Allgemeines Gleichbehandlungsgesetz (AGG)* (it: Legge Generale sulla parità di trattamento).

¹⁹ Lo studio ha analizzato testi provenienti dall'Agenzia di Stampa Tedesca (Deutsche Presseagentur) e dalle riviste *Brigitte*, *Zeit Wissen* e *Psychologie Heute*, tutti contenuti nel corpus DeReKo. I testi sono stati annotati manualmente per evidenziare quali elementi andrebbero cambiati per rendere il testo inclusivo dal punto di vista del genere.

epiceni come ad esempio *Person* (persona) o *Mensch* (essere umano), che consentono di riferirsi a persone senza specificarne il genere. Ciononostante, le critiche e l'ostilità verso il linguaggio di genere rimangono all'ordine del giorno.

Sono diversi i fattori che complicano la diffusione del linguaggio di genere e ostacolano il cambiamento linguistico in questo senso. Tra questi, un ruolo importante gioca il fattore cognitivo, legato all'abitudine. Utilizzare un linguaggio inclusivo, per chi non è abituato a farlo, non è automatico: è necessario fermarsi prima di parlare e riflettere consapevolmente su cosa vogliamo dire e soprattutto come vogliamo dirlo, il che richiede uno sforzo cognitivo maggiore rispetto al ricorso al linguaggio sessista (Waldendorf, 2024:360). Altri elementi sfavorevoli alla diffusione del linguaggio di genere sono legati al clima di antagonismo che il cambiamento linguistico in generale, e il linguaggio di genere in particolare, devono fronteggiare, nonché alla presenza di troppe varianti possibili, che possono disorientare e paralizzare l'oratore nelle sue scelte linguistiche (ivi, pp. 360-361). A queste problematiche si aggiungono la disomogeneità della terminologia utilizzata per categorizzare le varie proposte di linguaggio di genere e il fatto che spesso l'opinione pubblica si lascia influenzare da informazioni che talvolta non sono verificate o verificabili.

Secondo la psicolinguistica, l'atteggiamento di un/a parlante nei confronti del linguaggio di genere e il suo impiego concreto sono legati a una serie di fattori: tra questi, le convinzioni personali (Burnett & Pozniak, 2021; Formanowicz et al., 2013; Gustafsson Sendén et al, 2015; Sarrasin et al, 2012), l'interesse e la consapevolezza nei confronti dell'argomento (Gustafsson Sendén et al., 2015 & 2021; Judith, 2025)²⁰ e la dimestichezza con il linguaggio di genere (Bordin, 2022). È possibile dunque argomentare che le persone che mostrano una maggiore sensibilità e consapevolezza di genere, una maggiore dimestichezza con le strategie di genere e un maggiore interesse nella tematica saranno più propense a impiegare il linguaggio di genere nel proprio eloquio.

Curiosi sono poi, dal lato opposto dello spettro comunicativo, alcuni casi di eccesso di zelo, come nel caso della formula „*Mitgliederinnen und Mitglieder*“. Il *Duden* ci ricorda infatti che la parola *Mitglied* non può realizzare il suo femminile in -in tramite *Movierung*, perché *Mitglied* è neutro (*das Mitglied*) e fa riferimento sia a donne che a uomini. Di conseguenza, „*Liebe Mitglieder*“ va già bene così (Duden, s.d.). Lo spiegava già Luise Pusch, negli anni Ottanta:

²⁰ Per maggiori approfondimenti sui fattori correlati a posizionamento nei confronti del linguaggio di genere e impiego dello stesso si rimanda a Lange & Von Stockhausen (2025).

»DAS Mitglied«, dozierte ich, »ist doch eine der wenigen nun wirklich geschlechtsneutralen Personenbezeichnungen, die wir haben!²¹

(Pusch, 1984: 163)

1.7 Le principali soluzioni per l'impiego di un linguaggio inclusivo in tedesco

L'applicazione del linguaggio di genere in tedesco può risultare complicata per via della mancata standardizzazione e degli ostacoli insiti nel sistema linguistico. Non esiste una terminologia unitaria, e spesso gli stessi fenomeni vengono chiamati in maniera differente oppure si utilizza lo stesso termine per definire due fenomeni diversi.²²

Inoltre, diverse istituzioni, enti e organi seguono linee guida differenti, per cui ciò che è ammesso o anche consigliato in un dato contesto può essere inammissibile o scorretto in un altro. Mentre diverse università tedesche, come l'Università di Colonia (2021) e l'Università Heinrich-Heine di Düsseldorf (2023) consentono o persino incoraggiano l'uso di simboli grafici come l'asterisco e il trattino basso, altrettanti organi, enti e linee guida ne sconsigliano o addirittura ne vietano l'uso. È il caso, ad esempio, della Baviera, che tramite un atto normativo del 2024 ha stabilito che per le autorità amministrative bavaresi l'uso di questi simboli grafici non è consentito (§ 22 Allgemeine Geschäftsordnung für die Behörden des Freistaates Bayern (AGO)).

Davanti all'oceano di consigli, linee guida e manuali che talvolta si contraddicono tra loro è facile sentirsi disorientati, soprattutto per chi si approccia per la prima volta al *Gendern*, cioè all'impiego del linguaggio di genere in tedesco.

Per questo motivo ci proponiamo nella presente sezione di riassumere in due tabelle per sommi capi le principali soluzioni proposte dalle linee guida per il linguaggio di genere in tedesco, soffermandoci anche sulla terminologia che utilizzeremo nella presente tesi, al fine di garantire il più possibile uniformità e chiarezza.

Le tabelle si rifanno alle schematizzazioni contenute nella tesi magistrale di Sorbelli (2024) e nella tesi di dottorato di Nardone (2018), nonché alle linee guida fornite dall'Università di

²¹ It: ««DAS Mitglied», spiegavo io, «è uno dei pochi termini riferiti a persone realmente neutri che abbiamo!»»

²² La menzione di donne e uomini in riferimento a gruppi misti (a cui noi ci riferiamo con il termine sdoppiamento), prende ad esempio nomi diversi, come *Paarformen* (cfr. Cancelleria federale svizzera, 2023; Elmiger et al., 2017), *Beidnennung* o *Doppelnennung*, *Dopplung* o anche *splitting* (cfr. Dehoff-Zuch, 2016; Universität zu Köln, 2021; Gendern, n.d.; Pusch, 1984) in diversi contesti.

Lipsia e consultabili nella sezione “Gendern” del *Portal Wissenschaftliches Schreiben* (it: “portale per la scrittura accademica”) sul sito dell’Università (Gendern, n.d.) e alle linee guida pubblicate dalla Cancelleria federale svizzera nel 2023.

Nella colonna “Strategia” vengono elencate le diverse soluzioni proposte dalle varie linee guida; nella colonna “Tipologia” viene indicato se si tratta di strategie di neutralizzazione (N) o di visibilità (V). Inoltre, viene indicato se le strategie di visibilità rappresentano solo uomini e donne (VMF) o se includono anche persone non-binarie (VNB). Nella colonna “Esempi” vengono riportati alcuni esempi d’uso, mentre la colonna “commenti” raccoglie alcune osservazioni su lati positivi e negativi delle rispettive soluzioni.

Tabella 1 – Proposte di ordine terminologico²³

Strategia	Tipologia	Esempi	Commenti
Uso di aggettivi sostantivati al plurale	N	<i>Die Armen, die Kranken</i>	Questa soluzione rappresenta tutti i generi in un unico termine. Tuttavia, la soluzione è inclusiva solo al plurale, in quanto al singolare è richiesta l’esplicitazione del genere.
Uso di participi sostantivati al plurale	N	<i>Die Angestellten, die Mitarbeitende</i>	Questa soluzione rappresenta tutti i generi in un unico termine. Tuttavia, la soluzione è inclusiva solo al plurale, in quanto al singolare è richiesta l’esplicitazione del genere. Inoltre, non sempre i participi hanno lo stesso significato del termine che sostituiscono (ad es. <i>Fahrende</i> vs. <i>Fahrerinnen und Fahrer</i>).
Ricorso a sostantivi epiceni e suffissi neutri	N	<i>Person, Mensch, Individuum, Fachkraft, Bergleute, Bezugsperson</i>	Questa soluzione rappresenta tutti i generi in un unico termine. Alcuni sostantivi epiceni possono essere utilizzati all’interno di composti come dei “suffissi neutri” in composti come <i>Bezugsperson</i> o <i>Fachkraft</i> .

²³ La suddivisione in proposte di ordine terminologico e sintattico è ripresa da Sorbelli (2024).

Sostantivi collettivi	N	<i>Die Delegation, das Publikum, das Ministerium, die Wählerschaft, die Vertretung</i>	Anche questa soluzione rappresenta tutti i generi in un unico termine, ma non sempre il sostantivo collettivo ha esattamente lo stesso significato del sostantivo che specifica il genere (<i>die Delegation</i> ad esempio non equivale sempre <i>der/die Delegierte</i>). Inoltre, i collettivi spostano l'attenzione dal singolo alla collettività/al gruppo, mettendo in secondo piano l'individuo che compie l'azione. ²⁴
Sdoppiamento (per esteso)	VFM	<i>Schülerinnen und Schüler, Bäuerinnen und Bauer, Wählerinnen und Wähler</i>	Formulazione corretta dal punto di vista grammaticale, che rappresenta sia uomini che donne, ma non include le persone non binarie. Inoltre, comporta un allungamento della frase, che può incidere sulla leggibilità. ²⁵

²⁴ (Cancelleria federale svizzera, 2023:11).

²⁵ Da un lato, diverse linee guida sottolineano che un eccesso di forme sdoppiate può intaccare la leggibilità di un testo (Cancelleria federale svizzera, 2023:5). Dall'altro, è stato provato che forme di linguaggio inclusivo come lo sdoppiamento non inficiano né la comprensibilità di un testo, né tantomeno la capacità di memorizzazione del suo contenuto (Blake & Klimmt, 2010; Braun et al., 2007; Friedrich & Heise, 2019). È stato inoltre riscontrato che lo sdoppiamento porta a maggiori livelli di autostima: davanti a professioni al maschile generico, bambini e bambine le considerano generalmente più difficili da ottenere, a prescindere da genere, età e lingua madre, mentre si sentono più sicuri se confrontati con forme "sdoppiate" (Vervecken & Hannover 2015).

<i>Splitting</i> ²⁶	VFM	<i>Schüler/innen, Wähler/innen Hörer/innen</i>	Nello scritto è possibile “condensare” le forme sdoppiate in un unico termine tramite l’uso di uno slash prima del suffisso femminile in <i>-in</i> , evitando un allungamento eccessivo della frase ²⁷ . Questa soluzione non include persone non binarie, e va utilizzata con moderazione, perché può portare a formulazioni poco leggibili (ad es. <i>ein/-e neue/-r Mitarbeiter/-in</i>). Inoltre, sorgono problemi nello <i>splitting</i> di sostantivi che contengono un Umlaut (ad es. <i>Arzt oder Ärztin</i> non può essere abbreviato in <i>Ärzt/-in</i> , perché la forma maschile è <i>Arzt</i> e non <i>*Ärzt</i>).
Uso di <i>Binnen-I</i>	VFM	<i>SchülerInnen, WählerInnen; HörerInnen</i>	Questa soluzione rappresenta sia uomini che donne, ma non include persone non binarie. Inoltre, anche in questo caso sono problematiche le forme che contengono un <i>Umlaut</i> (ad es. <i>ÄrztIn</i> non è possibile, vedi sopra). La forma viene spesso criticata in quanto risulta errata secondo le regole ortografiche vigenti ²⁸ .

²⁶ Lo *splitting*, la *Binnen-I* e i segni grafici sono fenomeni della lingua scritta. Nel parlato, generalmente tutti questi fenomeni vengono pronunciati tramite un colpo di glottide. A questo fenomeno della lingua parlata viene dedicata una sezione a seguire (1.7.1).

²⁷ Nei testi scritti si sono anche stabilite abbreviazioni come *SuS* per la forma sdoppiata *Schülerinnen und Schüler*.

²⁸ Le linee guida della Cancelleria federale svizzera ad esempio vietano l’uso di *Binnen-I* e di segni grafici, affermando che „Sie entsprechen nicht der amtlichen deutschen Rechtschreibung, führen zu grammatisch falschen Formen und beeinträchtigen bei konsequenter Verwendung die Lesbarkeit von Texten“ (it: “Non corrispondono all’ortografia ufficiale tedesca, portano a forme grammaticalmente errate e, se utilizzate costantemente, compromettono la leggibilità dei testi” (2023:7), mentre le linee guida dell’Università di Lipsia affermano che „Die Form ist nach den gültigen Rechtschreibregeln eigentlich falsch, weil es innerhalb eines Wortes keine Großbuchstaben gibt. Sie lässt sich zudem schlecht beugen und noch schlechter lesen“ (it: “La forma è in realtà errata secondo le regole ortografiche vigenti, perché all’interno di una parola non possono esserci lettere maiuscole. Inoltre è difficile da declinare e ancora più difficile da leggere.”).

Uso di segni grafici	VNB	<i>Schüler*innen,</i> <i>Wähler:innen,</i> <i>Hörer_innen</i>	Forme come il <i>Gendersterchen</i> (*), il <i>Gender-Gap</i> () e i due punti (:) inserite tra la forma maschile e la desinenza femminile includono tutti i generi, comprese le persone non binarie, in un unico termine. Analogamente alla <i>Binnen-I</i> , anche quest'opzione viene criticata da diverse linee guida in quanto errata secondo le regole ortografiche vigenti e presenta le stesse problematiche per quanto riguarda le forme che contengono un <i>Umlaut</i> (ad es. <i>der*die Arzt*in</i> non è possibile).
-----------------------------	-----	---	---

Tabella 2 – Proposte di ordine sintattico

Strategia	Tipologia	Esempi	Commenti
Uso di pronomi indefiniti come <i>wer, alle</i> o <i>diejenige</i>	N	<i>Wer eine Zulassung beantragt, muss...</i> (invece di <i>der Antragsteller muss...</i>)	Viene neutralizzato il riferimento al genere. Tuttavia, può richiedere riformulazioni che appesantiscono il discorso. Inoltre, la ripresa di pronomi con <i>Genus</i> maschile come <i>wer</i> richiede l'esplicitazione del maschile (ad es.: "Wer das versteht, <u>der</u> kann es auch erklären.") e <i>derjenige/diejenige</i> al singolare richiede l'esplicitazione del genere.
Uso del passivo	N	<i>Es wurde herausgefunden, dass... (invece di Wissenschaftler haben herausgefunden, dass...)</i>	Soluzione che neutralizza i riferimenti al genere, ma vi è il rischio di appesantire e spersonalizzare il discorso.

Riformulazione impersonale	N	<i>Aus biologischer Sicht</i> (invece di <i>aus Sicht eines Biologen</i>); <i>ärztliche Hilfe</i> (invece di <i>die Hilfe eines Arztes</i>)	Vengono neutralizzati i riferimenti al genere. In generale, rende il discorso più ‘asettico’ e astratto.
Evitare l’uso di <i>man</i>²⁹	N	<i>Das kann auch so gesehen werden</i> (invece di <i>man kann das auch so sehen</i>)	Viene evitato l’uso del pronome <i>man</i> in quanto secondo alcuni richiamerebbe significati “maschili” in quanto grammaticalmente maschile (oltre che omofono del sostantivo <i>Mann</i> = uomo). Oltre alla riformulazione della frase, alcune linee guida consentono anche l’uso di pronomi alternativi, come <i>frau</i> , <i>eine</i> , <i>mann</i> ³⁰ e <i>mensch</i> .
Uso della forma di cortesia	N	<i>Bitte unterschreiben Sie das Formular</i> (invece di <i>Der Antragsteller muss das Formular unterschreiben</i>)	La forma di cortesia elimina il riferimento al genere, ma richiede di rivolgersi direttamente al soggetto, intaccando potenzialmente il registro. ³¹
Uso di aggettivi con sostantivi epiceni	N	<i>eine beratende Person</i> (invece di <i>Ein Berater</i>)	Viene evitato il riferimento al genere, ma l’aggiunta dell’aggettivo allunga e appesantisce la sintassi.

²⁹ Per quanto la sostituzione di *man* non sia tra le strategie di genere maggiormente consigliate dalle linee guida (Sorbelli, 2024:103), si tratta di una strategia riscontrata nel parlato dell’intervistata Luise Pusch all’interno di uno dei due corpora oggetto del presente studio (cfr. capitolo 3).

³⁰ È doveroso notare che nel parlato la forma *mann* non è utilizzabile perché si confonderebbe foneticamente con *man*.

³¹ Sorbelli (2024:139) afferma che questo tipo di riformulazione “richiede di rivolgersi direttamente alla persona interessata, con possibili conseguenze anche sul registro”.

1.7.1 Il colpo di glottide

Il colpo di glottide³² [ʔ] è un fonema presente in diverse lingue del mondo, spesso prima di parole che iniziano per vocale (Maddieson, 1984; Garellek, 2013, citato in Körner et al, 2024). Nel contesto del linguaggio di genere in tedesco, viene proposto come soluzione per la pronuncia delle forme inclusive che utilizzano *Binnen-I*, slash o segni grafici nella lingua scritta. Ha il vantaggio di essere un fonema già presente nella lingua tedesca, che quindi i parlanti di tedesco sono in grado di pronunciare (Haider, 2022, citato in Körner et al, 2024), e si realizza con una breve pausa tra una radice di parola maschile e una desinenza femminile. Permette di includere tutti i generi – anche identità di genere non binarie – in un unico termine.³³

Per quanto siano stati documentati esempi isolati dell'uso del colpo di glottide come indicatore di inclusione di uomini e donne già negli anni Ottanta, è soltanto intorno al 2019 che questo fonema comincia ad diffondersi nei media *mainstream* (Stephan, 2019; Fromm, 2020; citati in Körner et al., 2024). Trattandosi di un fenomeno relativamente recente, la ricerca sul colpo di glottide è limitata e talvolta contraddittoria: se da un lato Jöckel et al. (2021) riscontrano un calo di intelligibilità legato all'uso di questa forma, dall'altro Friedrich et al. (2022) non registrano alcun calo in questo senso.

Körner et al. (2024) hanno condotto uno studio interessante sul modo in cui il colpo di glottide viene percepito dalla popolazione, confrontandolo con forme di maschile generico e di femminile generico. Ne è risultato che il maschile generico ha portato a un *bias* maschile, mentre sia femminile generico che colpo di glottide hanno portato a un *bias* femminile. Nessuna delle tre forme prese in esame ha portato a rappresentazioni mentali ugualmente forti di donne e uomini, ma il colpo di glottide e il maschile generico hanno comportato *bias* di misura minore rispetto al femminile generico. Gli autori ne traggono la conseguenza che, per quanto l'esposizione al fenomeno sia stata limitata, il colpo di glottide porta già a rappresentazioni mentali di genere più inclusive rispetto al femminile generico.

³² Viene anche chiamato *Glottisschlag*, *Gender-Pause*, *Knacklaut* o *glottal stop*. Per garantire chiarezza terminologica nel presente studio ci riferiremo al fenomeno esclusivamente come colpo di glottide.

³³ Körner et al. (2024:159) definiscono il colpo di glottide come inclusivo di tutti i generi. Tuttavia, siccome può corrispondere nella forma scritta sia all'uso di *Binnen-I* (non inclusiva di persone non binarie) che all'uso di segni grafici (inclusivi di persone non binarie), risulta difficile disambiguare, nel parlato, se l'oratore/oratrice utilizza la forma in riferimento solo a donne e uomini o a tutti i generi.

CAPITOLO 2 – I CORPORA

Il presente capitolo descrive i metodi e i materiali adottati per condurre lo studio. Ad una breve panoramica sui *corpus studies* (2.1) e sull'impiego della linguistica dei corpora nello studio del linguaggio di genere (2.2) segue la descrizione delle peculiarità della creazione di un corpus di lingua parlata (2.3). Successivamente, viene descritto il processo che ha portato alla creazione del corpus *Alles Gesagt?* (2.4) e ci si sofferma sulle caratteristiche dei materiali che compongono il corpus e sul processo di trascrizione del parlato dialogico supportato da tecnologie di IA. Inoltre, vengono descritte le modalità di annotazione e interrogazione del corpus. Vengono poi illustrate le caratteristiche e i materiali che compongono il corpus *Politische Reden (1982-2020)* e le relative modalità di interrogazione (2.5). Infine, i due corpora vengono confrontati, evidenziando limiti e potenzialità dell'uno e dell'altro corpus (2.6).

2.1 La linguistica dei corpora

Il termine corpus (pl. corpora), deriva dal latino *corpus, corporis* (“corpo”). Fin dall'antichità classica, il termine è usato per indicare una raccolta di più opere di uno o più autori intorno a determinate materie: *Corpus Hippocraticum*; *Corpus iuris civilis*; *Corpus iuris canonici*; *Corpus inscriptionum latinarum* (Istituto dell'Enciclopedia Italiana, s.d.). Un corpus può raccogliere dati linguistici in formato elettronico e spesso corredati di strumenti di consultazione informatici (Baroni, 2010). Nel caso dei corpora di lingua parlata, o *spoken corpora*, il corpus può essere definito come “l'insieme dei discorsi registrati per una determinata analisi” (Fiehler, 2009:1187).

Oggi, il corpus rappresenta una risorsa straordinaria e un elemento chiave di molteplici studi linguistici, che permette di confrontare risultati ricavati da materiali più limitati con quantità di dati molto grandi, che altrimenti non sarebbero gestibili da una sola persona.

Secondo Bubenhofer (2009:16) lo studio di raccolte di testi non è nuovo nell'ambito della linguistica, ma la linguistica dei corpora moderna si distingue per via della grande mole di testo (un corpus può essere costituito da diverse migliaia di testi). Inoltre, nella linguistica dei corpora il focus non è rivolto all'uso della lingua nei singoli testi, bensì in grandi gruppi di testi, e si lavora in maniera empirica, con metodi quantitativi. Questo permette di considerare il linguaggio come una raccolta di dati in cui, con metodi induttivi, è possibile rendere visibili strutture statisticamente significative (ibid.).

I testi orali o scritti che compongono un corpus possono essere interrogati con specifici software per condurre vari tipi di analisi linguistiche, in base all'oggetto della ricerca. I testi sono generalmente corredati di metadati, ovvero informazioni riguardo all'autore, alla fonte o all'anno di provenienza del testo stesso. I corpora consentono di condurre indagini empiriche sulla lingua utilizzando tecniche automatiche e semiautomatiche, e di condurre sia un'analisi quantitativa su un grande numero di dati, che un'analisi qualitativa su un particolare fenomeno di interesse (Biber et al, 1998:4).

Attraverso l'uso dei computer, che facilitano l'analisi del linguaggio, è possibile identificare in modo rapido e accurato le regolarità tra centinaia o migliaia di testi (Baker & McEnery, 2015:1). In questo modo, diventa possibile analizzare una quantità di dati molto ampia più rapidamente e più accuratamente rispetto ad un'analisi manuale, ottenendo in modo automatico o semiautomatico una grande quantità di informazioni riguardo al corpus e alle sue componenti. La linguistica dei corpora si serve di vari strumenti per interrogare i corpora, applicando delle strategie di ricerca che permettono di individuare occorrenze (ovvero esempi o prove/riscontri) di un determinato fenomeno, per rilevarne la distribuzione all'interno del corpus (Bubenhofer 2013:109).

Teubert argomenta che la linguistica dei corpora non è di per sé un metodo, bensì esistono molti metodi diversi utilizzati nell'elaborazione e nell'analisi dei dati dei corpora. (Teubert, 2005:4). Tra le possibilità di analisi basate su corpora figurano ad esempio lo studio delle parole chiave, che evidenzia quali lessemi compaiono con maggiore frequenza in un corpus rispetto ad un corpus di riferimento, e il calcolo delle cooccorrenze o collocazioni, che permette di evidenziare in quali contesti compaiono determinati lessemi, ottenendo una lista di lessemi che compaiono più frequentemente insieme (Bubenhofer, 2013:109-116). Un'altra possibilità è rappresentata dal calcolo degli N-Grammi, ovvero unità composte da più parole (due, tre o più) adiacenti tra loro. Esistono anche N-Grammi complessi, composti da una combinazione di parole specifiche e di informazioni più generiche su parti del discorso (ad esempio *so+verbo finito+pronome personale*, che comprende N-Grammi del tipo “so machen wir”, “so steigen wir” etc.) (ivi, pp. 110-116).

Tali informazioni possono diventare il punto di partenza di analisi di vario genere. È possibile ad esempio verificare la diffusione di pattern individuati in corpus più piccolo tramite l'interrogazione di un corpus più grande, o anche confrontare i risultati ottenuti dalla stessa ricerca su corpora differenti. Tuttavia, è importante ricordare che l'obiettivo della linguistica

dei corpora non deve essere solo il riferire dati quantitativi, ma è fondamentale includere anche interpretazioni qualitative che permettano di esplorare l'importanza dei *pattern* quantitativi identificati (Biber et al., 1998:5).

Bubenhofers sostiene l'importanza di non orientarsi esclusivamente intorno alla frequenza, ma di adoperare bensì un'analisi qualitativa su base quantitativa (de: "quantitativ informierte qualitative Analyse"). Le analisi qualitative, in sostanza, devono essere condotte alla luce delle informazioni quantitative ottenute (Bubenhofers, 2013:109), e i dati qualitativi possono fornire spunti per analisi quantitative e viceversa.

2.2 Linguistica dei corpora e linguaggio di genere

La linguistica dei corpora è la metodologia adottata nell'ambito della presente tesi, dove viene applicata alla tematica del linguaggio di genere. Il linguaggio di genere in tedesco è stato a lungo una tematica poco ricercata nell'ambito della linguistica dei corpora (cfr. Haß-Zumkehr 2003; Storjohann 2004; Elmiger 2009; Elmiger 2011) mentre ad oggi gli studi che combinano il tema del genere con questa metodologia sembrano essere in graduale aumento (cfr. Elmiger et al., 2017a; Kopf, 2022; Kopf, 2023; Müller-Spitzer et al., 2024; Nübling & Rosar, 2025).

Ai fini di un migliore inquadramento teorico, ci sembra opportuno citare alcuni studi rilevanti, sviluppati in ambito germanista, in cui il tema del linguaggio di genere si interseca con la metodologia della linguistica dei corpora.

Nell'articolo *Hat die Frauenbewegung Wortschatzgeschichte geschrieben?* (2003) Haß-Zumkehr analizza una serie di fenomeni, tra cui la frequenza di sostantivi femminili terminanti in *-frau* e la presenza della *Binnen-I*, in due corpus dell'IDS³⁴ (Institut für Deutsche Sprache) Mannheim: il Corpus Storico (*Historisches Textkorpus*), che contiene documenti che vanno dal 1700 fino al 1945, e il corpus DeReKo, che contiene testi dal 1945 ai giorni nostri.³⁵ Storjohann (2004) ha utilizzato i corpora di lingua scritta dell'IDS Mannheim, per esplorare la diffusione e i contesti d'uso del pronome indefinito *frau*, confrontandola con le occorrenze delle alternative *man*, *mann* e *mensch*, osservando che l'uso del pronome *frau*, per quanto presente

³⁴ Risorse di corpora online dell'Istituto per la lingua tedesca di Leibniz (IDS, Institut für Deutsche Sprache): <https://www.ids-mannheim.de/onlineangebote/>.

³⁵ Tra i risultati dell'analisi diacronica figura un aumento della produttività dei sostantivi terminanti in *-frau*, mentre l'uso della *Binnen-I*, attestato nei corpora dell'IDS a partire dal 1987, sembra registrare un picco nel 1990 per poi calare nuovamente a partire dal 1995.

nei corpora, rimane sostanzialmente relegato a contesti legati a questioni di genere, linguistica femminista o linguaggio non sessista.

In *Sprachliche Gleichbehandlung von Frau und Mann: Eine korpusgestützte Untersuchung über den Sprachwandel in der Schweiz*, Elmiger (2009) ha utilizzato i due corpora *Schweizer Textkorpus*³⁶ e *W-öffentlich*³⁷ per analizzare quantitativamente la frequenza di diciotto forme femminili di sostantivi riferiti a persone (*Personenbezeichnungen*). Lo studio offre alcuni spunti interessanti, non solo sull'uso dei *Personenbezeichnungen* presi in analisi, ma anche su limiti e potenzialità dei due corpora utilizzati, nonché sulla possibilità di sviluppare, in futuro, corpora che combinino i punti di forza di entrambi.

Elmiger et al. (2017a) prendono in esame il linguaggio di genere all'interno del linguaggio amministrativo svizzero, combinando l'analisi *corpus-based* con altre metodologie: da un lato, viene preso in esame un corpus³⁸ che contiene tutti i testi pubblicati sul Foglio federale della Confederazione Svizzera (sul Foglio vengono pubblicati "in ordine cronologico tra l'altro rapporti del Consiglio federale, modifiche costituzionali o di legge e messaggi sulle votazioni popolari"³⁹) dal 1849 al 2014; dall'altro, vengono analizzati documenti normativi e vengono condotte interviste con funzionari, redattori, e traduttori dell'amministrazione pubblica svizzera, per scoprire come il linguaggio di genere viene concretamente impiegato. L'analisi dell'uso del linguaggio di genere nel corpus prende in esame alcune delle strategie di genere più comuni (participi ed aggettivi sostantivati, sostantivi collettivi, sostantivi epiceni, diverse forme di sdoppiamento per esteso e di *splitting*) e rivela che vengono utilizzate con maggiore frequenza lo sdoppiamento e gli epiceni come *Person* e *Mensch*. La triangolazione delle diverse metodologie permette di ottenere un quadro più completo, vario e sfaccettato.

Nübling e Rosar (2025) si concentrano invece sui *Personenbezeichnungen* in funzione predicativa dopo un soggetto femminile. Lo studio indaga il fenomeno della *Movierung*, o mancanza d'esso⁴⁰, all'interno del corpus *DeCoW*, e contestualmente riprende ricerche

³⁶ Lo *Schweizer Textkorpus* [...] è composto da circa 20 milioni di parole. Al momento dell'indagine (fine dicembre 2008/ inizio gennaio 2009) circa tre quarti di esse erano disponibili. (Elmiger, 2009:66).

³⁷ Il corpus *W-öffentlich* è composto da circa 1,1 miliardi di parole (dati aggiornati a fine dicembre 2008/inizio gennaio 2009). (Elmiger, 2009:66).

³⁸ Il corpus utilizzato nello studio si chiama *Bundesblatt / Feuille fédérale / Foglio federale* e raccoglie i contenuti della Cancelleria federale svizzera accessibili al pubblico in tedesco, francese e italiano, dal 1849 al 2014 (*Bundesblatt* e *Feuille fédérale*) e dal 1971 al 2014 (*Foglio federale*). Tuttavia, lo studio in sé si concentra principalmente sul corpus *Bundesblatt*, che contiene i documenti in tedesco.

³⁹ Fonte: Archivio federale svizzero (2025): <https://www.bar.admin.ch/bar/it/home/ricerca/ricerca/motori-di-ricerca-e-portali/pubblicazioni-ufficiali.html>.

⁴⁰ Un esempio della mancata *Movierung* a cui fanno riferimento Nübling e Rosar è la frase "Sie ist Arzt" (it. "Lei è un dottore"), in cui il sostantivo *Arzt* rimane al maschile nonostante il soggetto femminile, mentre la frase con

precedenti (cfr. Kopf, 2022; Kopf, 2023) su fattori considerati favorevoli o sfavorevoli al *Movierung* per verificarne l'effetto.

Riprendendo le categorie delle tabelle 1 e 2 (capitolo 1), è possibile osservare come alcune soluzioni, come ad esempio lo sdoppiamento, siano state ampiamente studiate nell'ambito della linguistica dei corpora (cfr. Waldendorf, 2024; Elmiger et al., 2017), mentre altre, soprattutto le soluzioni di ordine sintattico, sono state oggetto di meno studi. Lo sdoppiamento è forse tra le strategie di genere più ampiamente studiate, anche perché questa soluzione generalmente indica un uso specifico sia al singolare che al plurale, poiché contiene sia un sostantivo maschile che uno femminile (Elmiger et al., 2017b:33), mentre l'uso del passivo, della forma di cortesia o delle riformulazioni impersonali, per quanto contribuiscano all'inclusività del testo nell'insieme, non hanno come funzione primaria la sostituzione del maschile generico (Kolek, 2023:4).

Nell'ambito degli studi sul linguaggio di genere, i corpora possono essere impiegati per numerosi scopi, come analizzare il mutamento linguistico in prospettiva diacronica (cfr. Storjohann, 2004; Haß-Zumkehr 2003), indagare la diffusione e i contesti d'uso di diverse strategie di genere o anche valutare in modo empirico l'impatto di interventi di riforma linguistica. Inoltre, come mostrano gli approcci che integrano l'analisi *corpus-based* con altre metodologie, i corpora possono essere integrati in disegni di ricerca triangolati, capaci di restituire un quadro più complesso e sfaccettato delle pratiche linguistiche (cfr. Elmiger et al, 2017a). In questo senso, i corpora non solo consentono analisi sistematiche e replicabili, ma offrono anche prospettive promettenti di sviluppo futuro, ad esempio mediante la progettazione di nuovi corpora mirati.

2.3 La creazione di un corpus orale

Lo studio oggetto di questa tesi si concentra principalmente su aspetti legati al linguaggio di genere nel tedesco parlato. Entrambi i corpora analizzati sono infatti il risultato della trascrizione di parlato e rientrano quindi nella categoria dei corpus orali (cfr. Fiehler, 2009:1187).

Movierung sarebbe “Sie ist Ärztin”, con il sostantivo femminile *Ärztin* in funzione predicativa. Dallo studio emerge che la *Movierung* di *Personenbezeichnungen* riferiti a donne in posizione predicativa costituisce la norma (80-100%) mentre la mancata *Movierung* si verifica maggiormente in caso di prestiti dall'inglese o di forte genderizzazione maschile del concetto (es. *Soldat*, *Offizier*).

Tra i fattori che hanno determinato il focus sul tedesco parlato figura l'evidente prevalenza, quantomeno in ambito germanofono, di studi che si sono concentrati sul linguaggio di genere nel tedesco scritto (Hodel et al., 2017 & 2025; Ivanov & Lieboldt, 2023; Link, 2024; Müller-Spitzer et al., 2024; Pabst & Kollmayer, 2023; Preseau, 2025; Waldendorf, 2023 per citarne alcuni), mentre gli studi sul tedesco parlato sono meno numerosi, per quanto recentemente il focus sull'oralità sembra crescere (Körner et al., 2024; Chalyvidou & Weber, 2025; Bast et al., 2024). Questo sbilanciamento non è dovuto ad una mancanza di interesse o di rilevanza dell'argomento, ma piuttosto alle difficoltà aggiuntive che si presentano nella realizzazione e analisi di corpus orali, tra cui la necessità di trascrivere il parlato affinché possa essere studiato. La creazione e lo studio di un corpus di lingua parlata sono infatti compiti tutt'altro che semplici. A questo proposito, è sempre bene ricordare che le parole che compaiono all'interno di un tale corpus sono soltanto una rappresentazione parziale dell'evento comunicativo reale, mentre in un corpus di lingua scritta le parole esistono nella loro forma ortografica anche prima della creazione del corpus stesso (Straniero Sergio & Falbo, 2012:31).

2.3.1 Il 'written language bias' e il divario nella ricerca su tedesco scritto e orale

Le differenze tra il tedesco scritto e il tedesco parlato furono evidenziate già nel 1899 da Otto Behaghel (1927) in un suo discorso intitolato *Geschriebenes Deutsch und gesprochenes Deutsch* (it: "tedesco scritto e tedesco parlato") tenuto nell'ottobre del 1899, in occasione dell'assemblea generale del *Deutscher Sprachverein* ("Associazione linguistica tedesca"). Si era soffermato su caratteristiche come la melodia della frase, le variazioni di volume della voce e su fenomeni di accentuazione, e aveva avanzato alcune riflessioni su aspetti che oggi chiameremmo "non-verbali", nonché aspetti pragmatici che molto tempo dopo sarebbero stati focalizzati dall'analisi conversazionale:

Die Verschränkung des Gesprochenen mit dem, wie wir heute sagen, Nonverbalen, die Implizitheit und Knappheit der Äußerungen im Vergleich mit dem vom Sprecher Gemeinten und auch vom Hörer Verstandenen [...] – all das hatte Behagel gut beobachtet [...].⁴¹

(Schwitalla, 2003:14)

⁴¹ It: "L'intreccio tra il parlato e ciò che oggi chiamiamo non verbale, l'implicitezza e la sinteticità delle espressioni rispetto a ciò che l'oratore intende dire e che l'ascoltatore comprende [...]: Behagel aveva osservato bene tutto questo."

A partire dagli anni Trenta vennero sviluppate nuove tecnologie per la registrazione del suono, come il registratore a nastro. Man mano queste tecnologie divennero sempre più avanzate e soprattutto sempre più accessibili, permettendo uno studio più sistematico.

Già negli anni Settanta e Ottanta, gli studi di Rehbein ed Ehlich relativamente all'analisi dell'interazione e lo sviluppo del sistema di trascrizione HIAT (*Halbinterpretative Arbeitstranskriptionen*, Ehlich & Rehbein, 1976) avevano posto importanti basi metodologiche e terminologiche per la descrizione del parlato e per lo studio delle interazioni. Ulteriori contributi significativi sono la raccolta *Syntax des gesprochenen Deutsch* di Peter Schlobinski (1997), l'introduzione alla lingua parlata di Schwitalla (1997), la *Grammatik der gesprochenen Sprache* di Ludger Hoffmann (1998) e la successiva grammatica della lingua parlata di Mathilde Hennig (2008).

Nel capitolo *Gesprochene Sprache* dell'ottava edizione della Duden-Grammatik (2009), Fiehler riflette sulle differenze tra l'analisi dello scritto e dell'orale in tedesco, affermando che:

Die auf die Schriftsprache ausgerichteten Analyse- und Beschreibungskategorien sind darüber hinaus das einzige voll entwickelte Kategoriensystem. Ein Kategoriensystem, das in ähnlicher Weise funktional auf die gesprochene Sprache zugeschnitten wäre, existiert im Moment nur in Grundzügen.⁴²

(Fiehler, 2009:1168)

La lingua orale viene spesso infatti analizzata, interpretata e categorizzata secondo il modello di quella scritta, e la grammatica tradizionale è sostanzialmente e in prima linea una grammatica dello scritto (ivi, p. 1166).

Linell (1982) a tale proposito parla di un 'written language bias', dovuto al fatto che nelle società con una lingua scritta sviluppata la consapevolezza linguistica si basa appunto sulla lingua scritta. Il 'written language bias' condiziona anche la ricerca linguistica (Fiehler, 2009:1167). Ne consegue che i testi scritti sono l'oggetto di ricerca prevalente, e che gli strumenti di ricerca prevalenti sono categorie di analisi sviluppate durante e per l'analisi di testi scritti (ibid.).

⁴² It: "Le categorie di analisi e descrizione orientate alla lingua scritta sono inoltre l'unico sistema di categorizzazione pienamente sviluppato. Un sistema di categorizzazione che fosse funzionalmente adattato in modo simile alla lingua parlata esiste attualmente solo in forma embrionale."

2.3.2 Caratteristiche della comunicazione orale

Un fattore rilevante nel divario tra scritto e orale che ha a lungo rallentato la ricerca sull'oralità è dovuto ad una caratteristica intrinseca del parlato, ovvero la sua **fugacità**. Mentre i testi scritti sono sempre disponibili e possono essere analizzati ripetutamente e a lungo, i discorsi orali, per poter essere studiati, devono essere necessariamente trasformati in testi scritti (Fiehler, 2009:1168). Con lo sviluppo e la diffusione di appositi strumenti per la registrazione, la conservazione e la riproduzione di conversazioni e interazioni (videocamere, registratori a nastro e a cassetta) nonché con il recente e rapido sviluppo di tecnologie di *automatic speech recognition* (ASR)⁴³, il compito della trascrizione è diventato meno gravoso.

Un'altra caratteristica che distingue l'orale dallo scritto è la sua marcata **multimodalità**. Quando i/le parlanti interagiscono in copresenza, la comunicazione avviene simultaneamente e parallelamente su diversi livelli: la comunicazione fisica (linguaggio del corpo) e le informazioni basate su percezioni visive e conclusioni viaggiano sul canale visivo⁴⁴, mentre la comunicazione verbale orale (il parlato) viaggia sul canale acustico. La comunicazione orale coinvolge dunque contemporaneamente diversi sensi, mentre la comunicazione scritta è esclusivamente verbale e avviene solo visivamente (Fiehler, 2009:1189-1190).

Tra gli aspetti che caratterizzano l'interazione orale, Fiehler (ivi, p. 1179) sottolinea inoltre la copresenza delle parti e la condivisione della stessa situazione, la reciprocità della percezione, e l'interattività.

La lingua orale è strettamente legata al contesto comunicativo in cui avviene, e presenta alcune limitazioni dal punto di vista cognitivo, legate sia alla capacità di **pianificare** e ricordare il messaggio per chi parla, sia alla capacità di **elaborazione** del messaggio da parte di chi ascolta (Imo & Lanwer, 2016:156). Un testo scritto può essere riscritto e riletto più volte, e se ne può interrompere la fruizione, spostandosi liberamente all'interno del testo con i propri ritmi, aspetto che agevola sia la produzione che la ricezione di frasi particolarmente complesse; la ricezione dell'oralità deve avvenire invece in tempo reale e non consente sospensioni o passi indietro. Per questo motivo, quando si formula liberamente e sul momento, si manifestano alcuni fenomeni tipici dell'oralità, come **lapsus**, **interruzioni**, **false partenze** e **correzioni** di ogni tipo. A causa dell'**irreversibilità** della comunicazione orale, tutti questi processi lasciano tracce negli enunciati (nel nostro caso, nelle registrazioni). Generalmente, quando viene

⁴³ Alla trascrizione supportata da tecnologie di ASR è dedicata più avanti un'apposita sezione (cfr. capitolo 2.4.3).

⁴⁴ Fatta eccezione per le telefonate o per le situazioni di comunicazione a distanza senza connessione video.

riprodotto, ovvero letto o recitato ad alta voce un discorso pre-formulato per iscritto, questi fenomeni tendono a verificarsi meno frequentemente, anche se non sono da escludere del tutto (Fiehler, 2009:1186).

Tutti questi aspetti sono rilevanti anche per l'uso del linguaggio di genere nel parlato e nell'ambito dell'interpretazione. L'irreversibilità del parlato e la necessità di formulare *on-line*, ovvero, per così dire, in tempo reale, comportano che l'oratore ha meno tempo a disposizione per decidere quale strategia di genere usare, e mentre chi scrive un testo può sperimentare, testando varie soluzioni fino a individuare quale effettivamente funziona meglio in un dato contesto, chi parla spesso non può permettersi questo lusso.

Per questo motivo, sia nella comunicazione quotidiana che nel parlato nel parlato pubblico e nell'interpretazione è fondamentale per chi parla avere sviluppato automatismi, strategie “di default” su cui poter contare, che permettono velocizzare la formulazione e la produzione nel momento del bisogno. Attraverso l'analisi dei due corpora ci si pone quindi anche l'obiettivo di offrire spunti per sviluppare strategie utili anche nel contesto interpretativo verso il tedesco, dove alla volontà di impiegare un linguaggio rispettoso dal punto di vista del genere si somma anche la necessità di gestire le complessità cognitive intrinseche della pratica del parlato in pubblico, sia esso spontaneo o interpretato.

2.4 La creazione del corpus *Alles Gesagt?*

2.4.1 La scelta del materiale: Luise Pusch e il podcast *Alles Gesagt?*

Il primo passo è stato la scelta del materiale da utilizzare per la creazione del corpus. Si è deciso di realizzare un corpus ad hoc a partire da un episodio del podcast *Alles Gesagt?* in cui viene intervistata la linguista Luise Pusch⁴⁵, che è considerata una delle “madri fondatrici” della linguistica femminista tedesca, e ha dedicato la sua carriera alla valorizzazione delle donne e alla lotta contro il patriarcato.

Luise Pusch è nata nel 1944 a Gütersloh con il nome di Frohmut Pusch. Con il “cambiamento linguistico” ha incominciato già da giovanissima, abbandonando il suo nome di battesimo, che a suo dire “non le si addiceva” e cambiandolo in Luise.

⁴⁵ L'episodio, intitolato *Luise Pusch, warum ist Deutsch eine Männersprache?* è stato pubblicato dallo *Zeit* il 22 novembre 2022 e disponibile al link: <https://www.zeit.de/gesellschaft/2022-11/luise-pusch-gendern-deutschland-interviewpodcast-alles-gesagt> (consultato il 10.02.2026).

Nel 1979 Hartwig Kalverkämper pubblica *Die Frauen und die Sprache*, una critica in cui attacca duramente la linguista Trömel-Plötz, che in *Linguistik und Frauensprache* (1978) si era espressa infavore di una maggiore visibilità delle donne e contro l'uso del maschile generico, proponendo lo studio del 'linguaggio femminile' come legittimo oggetto di ricerca. Nello stesso anno, Luise Pusch risponde alle critiche di Kalverkämper con *Der Mensch ist ein Gewohnheitstier, doch weiter kommt man ohne ihr: Eine Antwort auf Kalverkämpers Kritik an Trömel-Plötz*, in cui argomenta la necessità di una critica linguistica in chiave femminista (Pusch, 1979). La risposta di Pusch a Kalverkämper segna un momento di svolta, aprendo la strada ad una lunga e prolifica carriera dedicata al linguaggio di genere e alla linguistica femminista.

Tra le opere più conosciute della linguista figura il libro *Das Deutsche als Männersprache: Diagnose und Therapievorschlge* (it: Il tedesco come lingua maschile: diagnosi e proposte terapeutiche), pubblicato per la prima volta nel 1984, che  oggi considerato un classico della linguistica femminista.

Nel podcast infinito *Alles Gesagt?*, la linguista parla delle differenze tra le diverse generazioni di femminismo, delle sue ambizioni professionali e del suo rapporto con la fede, ma racconta anche molto delle sue passioni e della sua vita personale, e di come si pu lottare contro il maschilismo usando la lingua come arma.

La linguista riflette inoltre sulla sua difficile infanzia e turbolenta adolescenza⁴⁶ in quanto donna lesbica in un'epoca in cui essere gay era illegale, ma essere lesbica non lo era, perch "le donne erano sempre state cos poco importanti che articoli di legge specifici al riguardo non esistevano nemmeno".

Interviewer: „Wie war es frher, lesbisch zu sein?“

Luise Pusch: „Es war einfach unertrglich. Es war permanente Angst.“⁴⁷

(*Alles Gesagt?*, 2002, 00:28:09 - 00:28:15. Pusch)

Also nachtrglich frage ich mich eben oft, wie ich das berlebt habe.⁴⁸

⁴⁶ Pusch racconta questi difficili anni anche nel libro *Gegen das Schweigen: Meine etwas andere Kindheit und Jugend* (it: "Contro il silenzio: la mia infanzia e adolescenza un po' fuori dal comune") (Pusch, 2022).

⁴⁷ Intervistatore: «Com'era essere lesbica, al tempo?»

Luise Pusch: «Era semplicemente insopportabile. Era una paura costante».

⁴⁸ "Col senno di poi mi chiedo spesso come ho fatto a sopravvivere."

Oltre al suo lavoro linguistico, ha creato una banca dati biografica online che raccoglie le biografie di 30.000 donne.⁴⁹ Oggi, la linguista vive tra Hannover e Boston, e dall'anno scorso è sposata con la sua compagna di lunga data, la germanista americana Joey Horsley. Nel 2025 è stata insignita della croce dell'Ordine al Merito della Repubblica Federale di Germania (Pusch, 2025).

Alles Gesagt (it: “Detto tutto?”) è un podcast del settimanale tedesco *die Zeit*. Il podcast si contraddistingue per il suo format fuori dal comune: gli episodi di *Alles Gesagt?* hanno infatti una durata potenzialmente infinita, e terminano solo ed esclusivamente quanto l'intervistato/a ritiene di aver “detto tutto”. A quel punto, l'intervistato/a pronuncia una parola che ha scelto appositamente a tale fine all'inizio dell'episodio, e pone fine così fine all'intervista.

Gli intervistatori Christoph Amend, direttore editoriale dello *Zeit*, e Jochen Wegner, caporedattore, preparano per ogni episodio centinaia di domande e spesso anche alcuni giochi per rendere più varia la conversazione. L'inusuale format del podcast produce spesso episodi molto lunghi, come nel caso dell'intervista allo youtuber Rezo, che è durata più di otto ore e mezza.

Le motivazioni che hanno portato alla scelta di questo podcast e in particolare di quest'intervista come base per il corpus sono principalmente di natura pratica (1), tecnica (2) e metodologica (3), ma tutte importanti per la buona riuscita del progetto. L'episodio presenta infatti una serie di caratteristiche favorevoli all'analisi che si desidera condurre:

- (1) L'episodio dura 5 ore e 23 minuti, fornendo molto materiale per la creazione del corpus;

una delle caratteristiche che ci hanno spinto ad esplorare questo podcast è proprio la lunghezza degli episodi e di conseguenza la grande mole di parlato di una medesima persona, oltre al fatto che lo *Zeit* lo metta a disposizione online senza particolari difficoltà di accesso.

- (2) La qualità dell'audio e le condizioni di registrazione sono favorevoli per l'elaborazione con software di trascrizione automatica;

⁴⁹ Consultabile sul sito FemBio (Frauen-Biographieforschung) al link <http://www.fembio.org/biographie.php/> (Recuperato il 12 Gennaio 2026).

gli episodi, da quello che è possibile dedurre dalla qualità delle puntate messe online, vengono registrati con attrezzature che permettono una buona qualità del suono. Il parlato è generalmente monolingue, lineare e privo di interruzioni (salvo brevi jingle o piccoli intermezzi pubblicitari). L'atmosfera è informale e di conseguenza anche il registro è informale, ma sempre curato, da persone abituate a parlare in e per un pubblico.

- (3) L'intervistata è sensibile alla causa del linguaggio di genere ed è solita impiegare un linguaggio rispettoso dei generi e consapevole, ci si aspetta dunque che impieghi strategie di genere anche in questa situazione comunicativa

Sulla base dei fattori che secondo la psicolinguistica influenzano l'impiego del linguaggio di genere da parte di un/a parlante (cfr. capitolo 1), è possibile argomentare che una persona come Luise Pusch, una figura di riferimento nell'ambito della linguistica femminista, che mostra una marcata sensibilità e consapevolezza di genere, nonché una grande dimestichezza con le strategie di genere e un forte interesse nella tematica – come si desume dai suoi scritti e da quanto lei stessa riferisce sulle circostanze della sua vita direttamente nell'intervista – sarà più propensa a impiegare il linguaggio di genere nel proprio eloquio, e di conseguenza di almeno alcune tra le strategie di genere elencate nel capitolo 1.

2.4.2 La fase di trascrizione

La fase di trascrizione ha come obiettivo la produzione di una traccia scritta di quanto detto che conservi al meglio le caratteristiche della conversazione o del discorso che si vuole analizzare, riflettendo il più fedelmente possibile in forma scritta le caratteristiche giudicate più importanti per l'analisi.

La trascrizione dell'oralità è solitamente un processo molto complesso che richiede una serie di scelte interpretative e una grande mole di lavoro: basti pensare che la quantità di tempo necessaria per realizzare una trascrizione può corrispondere a 30 o addirittura 60 volte la durata della registrazione (Fiehler, 2009:1187). Nel nostro caso, sfruttando tecnologie di trascrizione automatica è stato possibile ridurre significativamente questi tempi.

In base al tipo di analisi che si vuole effettuare, esistono diverse convenzioni che regolano come e cosa trascrivere. Un esempio di sistema di trascrizione è la trascrizione fonetica, che riproduce

in forma scritta i suoni di una lingua, ma esistono anche altri sistemi più vicini all'ortografia standard e maggiormente concentrati sul contenuto, ed altri ancora che permettono di evidenziare caratteristiche prosodiche del parlato, aspetti multimodali o legati al linguaggio del corpo. Un esempio di convenzione per le trascrizioni fonetiche è il sistema IPA, ovvero l'alfabeto fonetico internazionale, (ivi, p. 1188) mentre per la riproduzione di elementi non verbali e fenomeni prosodici vengono spesso impiegate le convenzioni di trascrizione del parlato e degli studi sull'interazione di Sacks, Schegloff e Jefferson (1974). Nell'ambito della germanistica è importante menzionare anche i sistemi di trascrizione HIAT (Ehlich & Rehbein, 1976)⁵⁰ e GAT 2 (Selting et al., 2009)⁵¹.

Nel nostro caso, abbiamo deciso di concentrarci sugli aspetti verbali (cosa viene detto), mettendo in secondo piano invece il piano non-verbale (come viene detto qualcosa)⁵². Questa decisione è stata determinata, oltre che da limitazioni di tempo, anche dalla consapevolezza del fatto che le strategie di genere che intendiamo esplorare si manifestano ad ogni modo prettamente sul piano verbale (menzione o mancata menzione del genere femminile, ricorso a participi o aggettivi sostantivati etc.).

Per contrassegnare la presenza di colpo di glottide nella trascrizione, il tool di trascrizione *Whisper* ha inserito uno spazio tra forma maschile e desinenza femminile (ad es. "Hörer innen"). Abbiamo deciso di mantenere questo tipo di contrassegno per il colpo di glottide piuttosto che contrassegnarlo con *slash*, *Binnen-I* o segni grafici. Questa decisione nasce dalla consapevolezza del fatto che tutte le principali varianti di trascrizione di questo fenomeno nella lingua scritta (*Binnen-I*, segni grafici, slash) hanno un significato ben preciso (ad esempio, la *Binnen-I* non include persone non binarie), mentre la soluzione adottata non "interpreta" il colpo di glottide dei parlanti e garantisce uniformità alla trascrizione.

2.4.3 La trascrizione supportata da riconoscimento automatico del parlato

Per ottenere una prima trascrizione grezza dei turni di parola dei partecipanti all'episodio ci siamo avvalsi di strumenti di trascrizione automatica, che ci hanno permesso di ridurre

⁵⁰ Per maggiori informazioni sul sistema HIAT è possibile consultare il seguente link: <https://exmaralda.org/en/hiat-2/>.

⁵¹ Per maggiori informazioni sul sistema GAT 2 è possibile consultare il seguente link: <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf>.

⁵² Per quanto comunque è doveroso notare che la possibilità di ascoltare e ri-ascoltare più volte la traccia audio originale ci permetta di avere qualche info aggiuntiva su come viene detto un determinato elemento, il che si rivelerà specialmente utile, ad esempio, nel disambiguare i casi di colpo di glottide dai semplici femminili plurali.

significativamente i tempi di trascrizione, pur dovendo comunque sottoporre la trascrizione grezza ad un'accurata revisione manuale prima di poterla utilizzare come base per il corpus.

La trascrizione automatica si basa sul riconoscimento automatico del parlato (in inglese ASR, *automatic speech recognition*), che permette di convertire il parlato in un testo scritto, attraverso un processo chiamato *speech-to-text-synthesis* (Calamai, 2023:2). Per ottenere le trascrizioni di parlato, è necessario impiegare un software appositamente 'addestrato' con una grande mole di dati, tra cui diverse ore di parlato trascritto manualmente. Il software può riconoscere solo ciò che già conosce, per cui maggiore e migliore è l'addestramento, e migliore sarà la capacità di riconoscimento (ibid.). Oggi, la ASR si basa principalmente sulle reti neurali artificiali profonde (il cosiddetto *deep learning*), che sono in grado di trascrivere segnali acustici direttamente in caratteri. (ibid.)

Questi sistemi tuttavia non sono infallibili, ma incontrano svariate difficoltà, ad es. in caso di rumore ambientale, parlato ipoarticolato e veloce, e all'aumentare delle dimensioni del vocabolario utilizzato dai parlanti. L'ASR lavora molto bene in ambiti come la dettatura in ambito giuridico o medico, dove raggiunge un tasso di errore (in inglese *word error rate*, l'acronimo inglese è WER) di 2-5%, ma se confrontato con conversazioni spontanee in contesti informali raggiunge tassi di errore del 40% (ivi, p. 3) – anche se sono percentuali difficilmente generalizzabili in quanto dipendono da un elevato numero di variabili.

The factors favourable to ASR: a single speaker, high quality recordings with very little background or technical noise, a formal style of speaking, e.g. reading or speech with no overlap, standard language, and a limited domain, i.e. a small vocabulary. [...] values that are detrimental to ASR performance: several speakers, echoes and distortions in the signal, informal way of speaking, i.e. a lot of overlapping speech and frequent interruptions, language with a strong accent or dialect, and unlimited domains, i.e. virtually unrestricted vocabularies.

(Draxler, 2022:35).

Nel nostro caso, il file presenta elementi favorevoli al lavoro con ASR:

- La traccia audio è stata evidentemente registrata con attrezzature e microfoni professionali, è di buona qualità, e presenta poco rumore di fondo;
- La lingua parlata è tedesco standard e non vengono pronunciate frasi in dialetto;

- Il parlato è quasi esclusivamente monolingue, ad eccezione di qualche sporadica citazione in inglese.

Al contempo, alcuni fattori della traccia audio, tutti strettamente legati alle caratteristiche che hanno portato alla selezione di quel testo (e pertanto inevitabili), hanno intralciato il lavoro di riconoscimento automatico:

- Il parlato non è monologico ma dialogico, con un numero elevato di parlanti e frequenti sovrapposizioni, interruzioni, correzioni e false partenze;

Come precedentemente accennato, dunque, prima di passare alla creazione del corpus e quindi alla fase di analisi, è stata necessaria una fase di revisione delle trascrizioni ottenute volta a correggere eventuali errori. Le trascrizioni sono state esportate in formato SRT per poi essere revisionate e corrette mediante un controllo concomitante con l'ascolto della traccia audio⁵³.

2.4.4 La trascrizione del parlato dialogico supportata da tecnologie IA

- Obiettivi del progetto e strumenti impiegati

La fase di trascrizione ha richiesto un grande lavoro di esplorazione e sperimentazione delle diverse possibilità di automazione del processo, in grado non solo di trascrivere correttamente un file di lunga durata, ma anche di identificare correttamente i/le vari/e parlanti, separando i loro turni di parola e producendo delle trascrizioni grezze che fossero già abbastanza precise, in modo da minimizzare i tempi di revisione e minimizzare l'intervento manuale, mantenendo al contempo alta qualità e precisione. La durata dell'episodio, la presenza di diversi oratori e le frequenti sovrapposizioni tra gli stessi hanno rappresentato delle complessità aggiuntive, che si è cercato di arginare implementando diverse soluzioni.

Per la realizzazione delle trascrizioni ci siamo avvalsi del supporto di Gabriele Carioli, programmatore, sistemista e tutor didattico presso il Dipartimento di Interpretazione e Traduzione di Forlì, che ha ideato e realizzato un software appositamente pensato per andare

⁵³ La fase di revisione delle 5 ore e 23 minuti di parlato trascritto ha richiesto circa una settimana di lavoro. Dal punto di vista semantico, le trascrizioni grezze ottenute contenevano pochi errori, perlopiù concentrati nelle sezioni ricche di sovrapposizioni oppure in rari segmenti che presentavano allucinazioni o turni di parola mancanti. Dal punto di vista formale, sono stati corretti alcuni errori ortografici, soprattutto dovuti alla presenza di nomi propri non noti al software. Per quanto riguarda l'attribuzione dei turni di parola, invece, il software di *diarizzazione* ha avuto maggiori difficoltà a distinguere i turni dei due intervistatori, probabilmente a causa della somiglianza nel timbro vocale, mentre ha commesso pochi errori nell'individuare la voce dell'intervistata (con maggiori errori in corrispondenza di sovrapposizioni o rapidi scambi di turno di parola).

incontro alle specifiche necessità del progetto. Per maggiori dettagli sulle scelte tecnologiche effettuate per la realizzazione del software e sulla *pipeline* messa a punto da Carioli rimandiamo alla dettagliata descrizione da parte dell'autore stesso riportata in appendice sulla quale del resto si appoggia contenutisticamente la descrizione contenuta nella presente sezione.

Come prima cosa sono quindi stati individuati gli strumenti che ci avrebbero permesso di *diarizzare*, ovvero di identificare correttamente gli oratori, e di trascriverne i rispettivi turni di parola.

Oggi esistono molti modelli linguistici (LM) adatti a questo scopo, addestrati su decine di migliaia di ore di parlato. Tuttavia, i più performanti sono spesso servizi commerciali chiusi, che girano su datacenter remoti dotati di gruppi di schede grafiche potenti (GPU) e sono accessibili tramite servizi a pagamento (API). Se da un lato questo permette a chiunque di usare una potenza di calcolo altrimenti inarrivabile per un dipartimento universitario o una piccola-media impresa, dall'altro comporta costi non trascurabili e nessuna garanzia sulla futura disponibilità del servizio o sulla stabilità delle sue interfacce di programmazione.

L'alternativa pratica è quindi utilizzare modelli open source. Per farlo in modo efficiente è però necessario un hardware dedicato, come workstation equipaggiate di GPU adeguate. Il nostro Dipartimento di Interpretazione e Traduzione dispone di alcune workstation dotate di GPU in grado di eseguire questi modelli.

Tra i LM open source disponibili al momento dell'avvio del progetto, *Whisper* e *PyAnnote* sono stati selezionati in quanto si distinguono per essere i più maturi ed efficaci. Entrambi vantano un'ottima documentazione, dimensioni compatibili con l'hardware a nostra disposizione e, soprattutto, sono gratuiti e distribuibili con licenza Open Source.

Whisper è un sistema di riconoscimento vocale automatico (ASR) sviluppato da OpenAI e addestrato su 680.000 ore di dati supervisionati multilingue e multiuso raccolti dal web. L'ampiezza e l'eterogeneità del dataset di addestramento contribuiscono a una buona robustezza del modello nella gestione di accenti differenti, rumori ambientali e terminologia specialistica. Il sistema offre capacità robuste di trascrizione in un centinaio di lingue, tra cui inglese, spagnolo, francese, tedesco, cinese, giapponese, arabo, hindi e molte altre. Inoltre, può gestire anche contenuti in più lingue e supporta la traduzione automatica verso l'inglese. La trascrizione Whisper AI è una tecnologia avanzata di conversione *speech-to-text*, in grado cioè

di convertire direttamente il parlato in testo scritto, con livelli di precisione e accuratezza che si avvicinano a quelli della trascrizione umana. (OpenAI, n.d.; WhisperAI, n.d.)⁵⁴

La sperimentazione nell'ambito del presente progetto ne ha confermato efficacia e precisione per quanto riguarda lessico, ortografia, punteggiatura e segmentazione in periodi del parlato spontaneo. Al contempo, è stato necessario far fronte ai punti deboli del motore di trascrizione, che lavora al meglio con blocchi di audio di lunghezza moderata e coerenti⁵⁵, peggiora con segmenti frammentati, composti solo da riempitivi o vocaboli non lessicali (o comunque troppo brevi per determinare un contesto) e tende ad avere “allucinazioni” nei segmenti eccessivamente lunghi.

Pyannote è un set di strumenti specializzato nella *diarizzazione*, un processo che permette di determinare chi sta parlando in un determinato punto di un flusso audio, identificando e separando i diversi oratori. Una *diarizzazione* affidabile era fondamentale ai fini della nostra ricerca, in quanto era necessario isolare i turni di parola dell'intervistata da quelli degli intervistatori, in modo da poter studiare separatamente il parlato dei singoli partecipanti, e in particolare quello di Pusch (pyannoteAI, n.d.).⁵⁶ *Pyannote*, nel nostro caso, ha mostrato dei limiti nella gestione di segmenti di audio lunghi⁵⁷, provocando la degradazione progressiva della qualità della *diarizzazione*.

- L'organizzazione del lavoro tramite *pipeline*

Una volta individuati gli strumenti per la *diarization* e per la trascrizione automatica del file audio, si è proceduto con i primi tentativi e le prime valutazioni sui risultati ottenuti.

Per questo progetto, si è deciso di evitare la creazione di un'applicazione monolitica, che avrebbe richiesto tempi di sviluppo molto più lunghi, e si è optato piuttosto per un'organizzazione del lavoro sotto forma di *pipeline*, cioè una sequenza di operazioni “semplici” successive. Questo approccio permette di intervenire manualmente tra uno step e

⁵⁴ Per maggiori informazioni su *Whisper* o per consultare il paper, la scheda del modello o il codice si rimanda alla pagina dedicata sul sito di *OpenAI* (<https://openai.com/it-IT/index/whisper/>) o alla sezione dedicata alle domande frequenti sul sito di *WhisperAI* (<https://whisperai.com/faq>).

⁵⁵ *Whisper* è stato addestrato su segmenti audio di massimo 30 secondi. Per audio più lunghi, talvolta il modello funziona bene, ma altre volte possono verificarsi derive e allucinazioni (Radford et al., 2022). Nel nostro caso, i segmenti elaborati da *Whisper* sono stati tutti più corti di 10 minuti.

⁵⁶ Per maggiori informazioni su *pyAnnote* si rimanda alla pagina dedicata, consultabile al seguente link: <https://docs.pyannote.ai/introduction>.

⁵⁷ Una lunghezza fissa non è specificabile, in quanto dipende dalle caratteristiche dell'audio. Nel nostro caso, la sperimentazione empirica con il vecchio modello di *pyAnnote* ha rivelato molti errori nell'attribuzione dei segmenti agli speaker nella traccia audio elaborata per intero, mentre gli errori diminuivano con blocchi da un'ora, mezz'ora e così via. Nell'ambito del presente progetto abbiamo optato per blocchi di 30 minuti che garantivano una buona performance, ma non è possibile presumere che la stessa durata sia ottimale in qualunque contesto.

l'altro della *pipeline*, qualora necessario, per apportare eventuali modifiche, e poi riprendere il processo con lo step successivo.

In questo capitolo, per ragioni di brevità, ci limitiamo a riassumere per sommi capi le fasi del procedimento messo a punto da Carioli che dal file audio originale permette di ottenere le trascrizioni automatiche del parlato suddiviso per speaker.

Innanzitutto, l'audio originale viene convertito in WAV 16kHz mono. Successivamente, viene diviso in blocchi di durata personalizzabile, che vengono processati indipendentemente da *PyAnnote*. In questa fase avviene la *diarizzazione*, cioè il riconoscimento degli speaker. In seguito, i singoli blocchi vengono riaggregati e viene effettuata una “pulizia” dei segmenti di parlato, volta ad eliminare quelli non significativi. I segmenti vengono successivamente trascritti da *Whisper*. Infine, vengono generati i file di trascrizione, in formato SRT e come file di testo con e senza *timestamp*; vengono generati sia file suddivisi per speaker, che contengono i turni di parola di un singolo oratore e sono utili alla creazione di subcorpora individuali, sia un file di trascrizione completo, che contiene tutti i turni di parola e corrisponde alla trascrizione dell'interazione nella sua interezza. Segue la fase di revisione manuale delle trascrizioni ottenute.

Nel nostro caso, è stato utilizzato lo strumento di sottotitolaggio *SubtitleEdit*, che permette di allineare la trascrizione con l'audio originale e di correggere eventuali errori man mano che si scorre la traccia, potendo confrontare in qualsiasi momento la trascrizione con la traccia audio originale. Una volta revisionate, le trascrizioni sono pronte per essere importate su *SketchEngine* e utilizzate come base per il corpus.

- Limiti e potenzialità

Nella realizzazione e nell'utilizzo della *pipeline* messa a punto nell'ambito della presente tesi si è dovuto fare i conti con alcune limitazioni, intrinseche dei LM utilizzati, tra cui la qualità delle trascrizioni dipendente dalla qualità audio originale, la difficoltà nel gestire speaker dalle caratteristiche vocali molto simili, e le sovrapposizioni tra voci, con musica o effetti sonori.

Queste limitazioni influiscono sulla qualità della *diarizzazione* e della trascrizione, rendendo necessari alcuni ulteriori interventi manuali. Ad ogni modo, i risultati ottenuti, estremamente incoraggianti, mettono in luce le nuove possibilità che le tecnologie di IA oggi offrono in termini di trascrizione del parlato: con le giuste accortezze sarà sempre più facile impiegarle nella trascrizione di discorsi monologici e dialogici, brevi o anche molto lunghi. Questo apre nuove piste e possibilità per gli studi sui corpora orali. Una frontiera futura potrebbe essere

rappresentata dall'impiego dell'IA non solo nella creazione, ma anche nell'interrogazione del corpus stesso, eventualmente anche in un'ottica di confronto con i risultati ottenuti dall'interrogazione umana. Quello che è certo è che il sistema adottato ha rappresentato nell'ambito del presente elaborato un ottimo compromesso tra automazione e precisione, permettendo di ottenere trascrizioni grezze di eccellente qualità, eliminando la necessità di trascrizione grezza manuale e riducendo significativamente i tempi di revisione.

2.4.5 Annotazione e interrogazione del corpus: Sketch Engine e GermanRFTagger

Per la compilazione e conseguente interrogazione del corpus si è scelto di utilizzare *Sketch Engine*, uno strumento di analisi testuale online che permette di analizzare corpora contenenti miliardi di parole ed è utilizzato da linguisti, lessicografi, traduttori, editori, università, agenzie di traduzione e istituti linguistici in tutto il mondo.

Sketch Engine consente la creazione di corpora a partire sia da testi reperiti online, sia da testi caricati manualmente in vari formati. Una volta selezionato il corpus su cui si vuole lavorare, è possibile consultarlo e interrogarlo grazie alle molteplici funzionalità che il *tool* mette a disposizione, come la ricerca delle concordanze, che consente di individuare rapidamente tutte le occorrenze di una determinata parola o sequenza di parole all'interno del corpus e di visualizzare anche il contesto in cui compaiono. Sketch Engine include inoltre strumenti per l'analisi degli N-Grammi, l'individuazione di collocazioni e di parole chiave, e per il confronto tra corpora (Sketch Engine, n.d.).⁵⁸ Il corpus *Alles Gesagt?* è stato caricato su *Sketch Engine* sia per intero, che sotto forma di due subcorpora: il subcorpus *Pusch*, che contiene i turni di parola dell'intervistata Luise Pusch e il subcorpus *Intervistatori*, contenente i turni di parola degli intervistatori Jochen Wegner e Cristoph Amend.

Un passaggio fondamentale nella creazione di un corpus è l'annotazione per parti del discorso (in inglese *part-of-speech-tagging* o *POS-tagging*). La procedura consiste nell'associare un *tag*, ovvero un etichetta, ad ogni parola del corpus, per indicarne la classe grammaticale. Per indicare che *taken* è un participio passato, ad esempio, è possibile annotarlo come segue: *taken_VNN*, dove il tag *VNN* indica il participio passato (Leech, 1997:2).

⁵⁸ Per maggiori informazioni su *Sketch Engine* è possibile consultare la pagina dedicata al seguente link: <https://www.sketchengine.eu/> nonché le sezioni "User Guide", "FAQ" e "What can Sketch Engine do?" (<https://www.sketchengine.eu/what-can-sketch-engine-do/>; <https://www.sketchengine.eu/frequently-asked-questions/>; <https://www.sketchengine.eu/guide/>).

L'annotazione può essere effettuata manualmente o automaticamente ed essere più o meno dettagliata (il tag può limitarsi all'informazione sulla parte del discorso o contenere anche informazioni morfologiche e grammaticali della parola, quali numero, genere, caso, tempo verbale). Il primo progetto di annotazione computerizzata fu effettuato sul *Brown Corpus*: ne risultò che il 77 per cento delle parole erano state annotate correttamente, mentre la restante percentuale fu corretta manualmente negli anni successivi (Leech:1997, 8). Oggi l'annotazione automatica ha fatto molti passi avanti e i risultati sono ovviamente di gran lunga migliori.

Esistono diversi software per l'annotazione automatica di un corpus, detti *tagger*, specifici per lingua. Il *tagger* di default per l'annotazione di testi in tedesco in Sketch Engine è *German RFTagger*⁵⁹.

È possibile interrogare un corpus tramite una ricerca semplice oppure optare per modalità più avanzate, come la ricerca tramite *Corpus Query Language*. Il *Corpus Query Language* (CQL) è un linguaggio di interrogazione di corpora che permette di ricercare modelli grammaticali o lessicali complessi o effettuare ricerche che non possono essere condotte utilizzando l'interfaccia utente standard.⁶⁰

Il linguaggio CQL è stato sviluppato dal gruppo *Corpora and Lexicons* dell'IMS dell'Università di Stoccarda a partire dall'inizio degli anni '90.⁶¹ Le *query*, o stringhe di ricerca in CQL possono contenere una combinazione di parole e *tag*, oppure essere composte esclusivamente da *tag* (Sketch Engine, n.d.).

2.5 Il corpus *Politische Reden* (1982-2020) del DWDS

Il corpus *Politische Reden* (1982-2020) è disponibile sul portale del DWDS (Digitales Wörterbuch der deutschen Sprache) dal 2019. Si tratta di una raccolta di trascrizioni di discorsi politici consultabile online e liberamente accessibile, che fa parte della raccolta di corpora *Gegenwartskorpora mit freiem Zugang* (it: “Corpora contemporanei ad accesso libero”).

Attualmente la raccolta di testi comprende contiene 15.240 discorsi (27 012 456 token) di oltre 200 figure politiche di rilievo provenienti da 6 diversi paesi e regioni, per un

⁵⁹ È possibile ottenere maggiori informazioni su *GermanRFTagger* nella sezione “German RFTagger part-of-speech tagset” del sito di *Sketch Engine*: <https://www.sketchengine.eu/germanrftaggerpartofspeechtagset/>.

⁶⁰ Per approfondimenti sul *Corpus Query Language* è possibile consultare la sezione “CQL basics” sul sito di *Sketch Engine* (<https://www.sketchengine.eu/documentation/cql-basics/>).

⁶¹ La versione utilizzata in *Sketch Engine* è un'estensione del linguaggio originale e presenta alcune variazioni.

totale di circa 27 milioni di parole, e viene costantemente ampliata, sia in termini di volume che di copertura temporale. Gran parte del corpus è costituita da discorsi tenuti dai titolari delle quattro più alte cariche di stato in Germania: i presidenti federali, i cancellieri federali, i presidenti del Bundestag e i ministri degli Esteri con mandati compresi tra il 1982 e il 2020; un'altra parte è costituita dai discorsi di personalità politiche influenti provenienti dall'Austria e dalla Svizzera, nonché, a titolo esemplificativo, dall'Alto Adige, dalla regione orientale del Belgio e dal Liechtenstein.

(DWDS, s.d.-a. Traduzione mia.)

I testi e i metadati sono stati estratti automaticamente. Per tutti i testi, sono disponibili i metadati relativi a titolo/i, relatore, data e fonte, mentre altri, come quelli relativi al luogo, o alle parole chiave, sono disponibili solo per alcuni testi. Per la tokenizzazione è stato usato *SoMaJo* e per il tagging delle parti del discorso è stato utilizzato il tagger *TreeTagger* in passato, mentre ora viene utilizzato il tagger *SoMeWeTa* (Barbaresi, 2018:794).

Per quanto riguarda le convenzioni di trascrizione, questo corpus utilizza una trascrizione vicina all'ortografia standard e non tiene conto di elementi non verbali e non prevede la possibilità di ascoltare i discorsi originali. Relativamente ai segni grafici, diversi discorsi riportano talvolta il trattino basso (es. Unionsbürger_innen), talvolta lo slash (es. Mitbürger/innen) e talvolta la *Binnen-I* (es. ArbeitnehmerInnen). Per quanto le registrazioni di molti discorsi siano disponibili in rete, il fatto che il corpus non offra la possibilità diretta di ascoltare la traccia audio originale rende complesso determinare se e quali di queste forme indichino la presenza o meno di un colpo di glottide; non è stato riscontrato l'uso dell'asterisco (*) e dei due punti (:).

È importante sottolineare che è anche possibile selezionare singoli oratori e svolgere dunque un'analisi sul parlato di una singola persona.

Il corpus *Politische Reden (1982-2020)* si presta particolarmente bene all'analisi della diffusione delle strategie di genere nel tempo anche per via dei decenni che abbraccia: la raccolta comincia infatti dagli anni Ottanta, quando la sensibilità alla tematica del linguaggio di genere in Germania non era ancora così sviluppata come oggi, ma cominciavano a diffondersi le proposte della linguistica femminista (avanzate già a partire dagli anni Settanta), veniva pubblicato *Das Deutsche als Männersprache* (1984) e i Verdi entravano per la prima volta nel Parlamento tedesco, contribuendo alla diffusione del linguaggio di genere nell'ambito politico (Stecker et al., 2021). Nei decenni successivi, il linguaggio di genere si è gradualmente affermato e si sono moltiplicate le pubblicazioni a riguardo, tra studi e linee guida (cfr. capitolo

1 della presente tesi), fino ad arrivare ai giorni nostri. Per quanto il corpus si fermi al 2020, escludendo quindi eventuali sviluppi degli ultimi 6 anni, ci permette comunque di osservare l'uso del linguaggio in un periodo storico in cui veniva introdotta la *terza opzione* (cfr. capitolo 1), e soluzioni come il colpo di glottide cominciavano a diffondersi nei media (Körner et al., 2024:159).

2.5.1 Interrogazione del corpus *Politische Reden (1982-2020)*

I corpora del DWDS possiedono un proprio linguaggio di ricerca realizzato tramite il software DDC e diverso da quello utilizzato in Sketch Engine.

Degna di nota e particolarmente utile ai fini del presente studio è la sezione della guida dedicata alla ricerca di alcuni fenomeni di *Gendern*.

2.6 I due corpora a confronto

Di seguito vengono riassunte le principali caratteristiche dei due corpora:

Tabella 3 – Principali caratteristiche dei due corpora

	<i>Alles Gesagt?</i>	<i>Politische Reden (1982-2020)</i>
Dimensioni (in token)	50185	27 012 456
Situazione comunicativa	Intervista con registrazione audio destinata alla pubblicazione come podcast	Discorsi politici in varie occasioni, dalle interrogazioni parlamentari ai discorsi celebrativi in occasione di anniversari, congressi, inaugurazioni e così via
Arco temporale di riferimento	22 Novembre 2022	1982-2020
Trascrizione del colpo di glottide	Spazio vuoto tra forma maschile e desinenza femminile	Slash, <i>Binnen-I</i> o trattino basso
Trascrizione di elementi non verbali	No	No
Numero di oratori	3	~200

CAPITOLO 3 – ANALISI DEI CORPORA

La presente sezione è dedicata all'analisi dell'impiego del linguaggio di genere all'interno del corpus *Alles Gesagt?* (3.1) e del corpus *Politische Reden (1982-2020)* (3.2). Per il corpus *Alles Gesagt?* vengono analizzati: il ricorso a pronomi alternativi a *man* (3.1), il ricorso a sdoppiamento e colpo di glottide (3.1.2), l'uso di aggettivi e participi sostantivati (3.1.3), l'uso di collettivi ed epiceni nonché la combinazione tra participi o aggettivi e sostantivi epiceni (3.1.4) e la ripresa anaforica di pronomi indefiniti con *Genus* maschile (3.1.5). All'interno del corpus *Politische Reden (1982-2020)* viene invece analizzata la frequenza d'uso, in termini assoluti e in chiave diacronica, l'uso dei pronomi *man* e *frau* (3.2.1) e di diverse forme di sdoppiamento in forma contratta (3.2.2), nonché delle forme maschili generiche e inclusive relativamente a tre coppie appositamente selezionate di *Personenbezeichnungen* (3.2.3). Infine, vengono discussi i principali risultati delle due analisi (3.3).

3.1 Analisi del corpus *Alles Gesagt?*

3.1.1 Sull'utilizzo di *man* e delle sue alternative

Ci proponiamo, in questa sezione, di analizzare l'uso dei pronomi *man*, *frau*, *eine* (femminile), *einer* (maschile), *keine* (femminile) e *keiner* (maschile) nel parlato di intervistata ed intervistatori, al fine di rilevare se e in che misura i tre parlanti utilizzano queste strategie di genere.

frau

Per la ricerca di *frau* è stata utilizzata la query [word="frau"], che individua soltanto le forme di *frau* scritte con la f minuscola, permettendo di escludere falsi positivi come il sostantivo *Frau*. Per individuare le forme di *frau* scritte con lettera maiuscola perché a inizio frase, è stata utilizzata la query [word="\."] [word="Frau"] [tag="V.*"] che identifica tutti i casi in cui *Frau* compare a inizio frase (subito dopo il punto) e seguita da un verbo. Usando word invece di lemma, si cerca solo la forma specifica *Frau* e si escludono falsi positivi come *Frauen*.⁶²

⁶²Il verbo subito dopo *frau* a inizio frase dovrebbe permettere di eliminare alcuni falsi positivi del tipo „, Frau Merkel ...“ o „, Frau Pusch ...“. Tuttavia, la ricerca ha prodotto 3 falsi positivi nel subcorpus degli intervistatori, dovuti proprio alla formula „Frau Pusch“. Questo fenomeno è verosimilmente dovuto a un problema di disambiguazione: la parola *Pusch* non viene riconosciuta come un cognome, bensì come un'occorrenza del verbo *puschen*.

Tutte le occorrenze rilevate sono contenute nel subcorpus dell'intervistata (gli intervistatori non hanno usato *frau*). Di seguito ne riportiamo un esempio:

Zum Beispiel in den Talkshows bei Anne Will, noch bis weit in die Nullerjahre, blickte **frau** um sich, sah nur Männer und Anne Will.

(*Alles Gesagt?*, 22.11.22, 00:59:49 – 01:00:08, Pusch.)

È interessante notare come in alcuni casi *frau* compaia in coppia con *man*. A questo fenomeno, che compare anche in altri esempi a seguire, è dedicata una sezione apposita più avanti.

man

Il pronome *man* non presenta omonimia con altre parti del discorso in tedesco, per cui è bastato utilizzare la query [lemma="man"], che individua sia le occorrenze di *man* scritto con lettera minuscola che di *Man* scritto con lettera maiuscola. Era ragionevole supporre che, siccome il testo dell'episodio contiene brevi frasi in inglese, ci fosse qualche falso positivo per via dell'uso del sostantivo inglese *man*, ma la verifica delle occorrenze non ha rivelato in realtà alcun falso positivo.

La tabella 4 riporta i dati relativi all'uso di *frau* e *man* nei due subcorpora, sia per quanto riguarda la frequenza assoluta (il totale delle occorrenze individuate), sia relativamente alla frequenza relativa nei rispettivi subcorpora calcolata in WpM⁶³ (*Words per Million*) al fine di consentire la confrontabilità dei dati.

Tabella 4 – Uso di *frau* e *man* nel subcorpus *Pusch* e nel subcorpus *Intervistatori*

	Frequenza assoluta in <i>Pusch</i>	Frequenza relativa in <i>Pusch</i> (WpM)	Frequenza assoluta in <i>Intervistatori</i>	Frequenza relativa in <i>Intervistatori</i> (WpM)
Totale <i>frau</i>	7	263,53	0	0,00
Totale <i>man</i>	107	4028,31	141	9204,26

eine, einer e keine, keiner

Per la ricerca di questi pronomi si è inizialmente tentato di individuare una soluzione che li distinguesse dalle parole articolo *ein* e *kein*, in modo da ridurre il numero di falsi positivi. La query elaborata a questo scopo, che permette di escludere le occorrenze di queste forme seguite

⁶³ Il calcolo Words Per Million (WPM), o frequenza normalizzata, determina la frequenza di una parola o evento specifico su un totale di un milione. La formula è: (Numero di occorrenze/Totale parole)×1.000.000.

da sostantivi o aggettivi declinati⁶⁴, non è stata tuttavia molto efficace, in quanto ha prodotto spesso diversi falsi positivi: nella maggior parte dei casi, questi falsi positivi erano composti da casi in cui le forme ricercate avevano funzione di articolo, ad esempio prima di avverbi (ad es. „Es gibt eigentlich eine ganz einfache Antwort darauf“)⁶⁵, o da casi in cui il pronome era preceduto da articolo (ad es. „**Der** eine kriegt dann mehr Gewicht.“). Si è dunque operata una verifica manuale di tutte le occorrenze delle varie forme. I risultati relativi ai due subcorpora sono poi stati riassunti nella tabella 5.

Tabella 5 – Frequenze delle varie forme nel dettaglio

Soluzione di genere	Frequenza assoluta in <i>Pusch</i>	Frequenza relativa in <i>Pusch</i> (WpM)	Frequenza assoluta in <i>Intervistatori</i>	Frequenza relativa in <i>Intervistatori</i> (WpM)
<i>frau</i>	6	225,88	0	0,00
<i>man</i>	101	3802,42	129	8420,91
<i>Man</i>	6	225,88	12	783,34
<i>Frau</i>	1	37,65	0	0,00
<i>eine</i>	2	75,29	0	0,00
<i>einer (f.)</i>	1	37,65	0	0,00
<i>einer (m.)</i>	0	0,00	0	0,00
<i>einem</i>	1	37,65	0	0,00
<i>einen</i>	0	0,00	0	0,00
<i>keine</i>	0	0,00	0	0,00
<i>keiner (f.)</i>	1	37,65	0	0,00
<i>keiner (m.)</i>	1	37,65	0	0,00
<i>keinem</i>	1	37,65	0	0,00
<i>keinen</i>	0	0,00	0	0,00
totale forme femminili	11	414,12	0	0, 00
totale forme maschili	110	4141,25	141	9204,25

Dai risultati dell'analisi si evince che, per quanto l'intervistata usi *frau*, *eine*, e *keine*, le forme maschili *man*, *einer* e *keiner* vengono comunque utilizzate in misura nettamente maggiore, in un rapporto 1:10. Gli intervistatori non utilizzano *frau*, *eine* o *keine*, bensì utilizzano esclusivamente *man*. Questo non permette di trarre alcuna conclusione di portata generale.

⁶⁴ La query prevede l'aggiunta di [tag!="N.*" & tag!="ADJ.*"] dopo la forma che si intende cercare.

⁶⁵ Gli esempi citati sono tratti dal corpus *Alles Gesagt?*. Sottolineature e grassetti sono miei.

Tuttavia, questa sezione ha evidenziato le complessità della ricerca in CQL di forme come *eine*, *keine*, *einer* e *keiner*. Le formule elaborate a questo scopo presentano alcuni limiti, resi evidenti dall'alta percentuale di falsi positivi. Per corpora di dimensioni ridotte come *Alles Gesagt?*, la via più sicura rimane la verifica manuale caso per caso di tutte le occorrenze delle varie forme. In un corpus di grandi dimensioni, dove la verifica caso per caso diventa impraticabile, lo studio di questi fenomeni rimane ancora una sfida, che potrebbe essere esplorata maggiormente nel dettaglio in future analisi. Inoltre, casi come il falso positivo “Frau Pusch” per la query [word="\."] [word="Frau"] [tag="V.*"] hanno evidenziato alcuni limiti del POS tagging quando viene posto dinanzi a parole polifunzionali.

***frau*, *eine* e *keine* vs *man*, *einer* e *keiner* in sdoppiamento e da soli**

Osservando da vicino i contesti d'uso si nota come l'intervistata utilizzi *frau*, *eine* e *keine* sia da soli che in combinazione con *man*, *einer* o *keiner*, e talvolta in combinazione:

Aber in meiner Zeit war das alles dermaßen mit Schande belegt, dass **frau und man** eigentlich niemandem was dazu erzählen konnte.

(*Alles Gesagt?*, 22.11.22, 00:29:50 – 00:30:06, Pusch.)

L'utilizzo di *frau* e *man* in combinazione rappresenta un'ulteriore variante della strategia di sdoppiamento, e non è un fenomeno nuovo, bensì era stato già osservato da Storjohann (2004)⁶⁶.

Dallo studio di Storjohann (2004) emergeva inoltre come *frau* venisse utilizzato quasi esclusivamente per riferirsi a persone di genere femminile come *agens* o *patients*. A questo proposito è possibile osservare che nel subcorpus *Pusch*, quando *frau* o *eine* compaiono da soli, si trovano spesso in contesti marcatamente femminili come quello del lesbismo o del femminismo. Inoltre, quasi tutti gli esempi possono essere riconducibili ad auto-riferimenti, cioè casi in cui la parlante parla di se stessa⁶⁷.

⁶⁶ Storjohann aveva notato, nel proprio corpus di riferimento, che *frau* compariva spesso accompagnato da *man* in fenomeni di *splitting* come “man(n) und frau” o “frau und man(n)”.

⁶⁷ A questo proposito, è interessante notare come l'oratrice utilizzi sia *frau* (es. „Das soll **frau** ja tun, besonders in meinem Alter. Bewegung ist wichtig.“) che *man* (es. „Aber das muss **einem** dann natürlich erst mal auffallen. Und mir ist es nicht aufgefallen von selber und den Frauen vielleicht auch nur so ansatzweise.“) per riferirsi a se stessa.

Tabella 6 – Contesti d’uso dei pronomi *frau* e *eine* nel subcorpus *Pusch*

Left	KWIC	Right	Contesto
Zum Beispiel in den Talkshows bei Anne Will, noch bis weit in die Nullerjahre, blickte	frau	um sich, sah nur Männer und Anne Will.	Femminismo
Also, dass	frau	vorsichtig mit Heterofrauen ist, mit Familienmüttern.	Lesbismo + Probabile auto-riferimento
Bei Lesben jedenfalls fällt das nicht auf einen Tag, sondern eher verteilt sich auf ein Jahr, bis	frau	dann zusammenzieht.	Lesbismo + Probabile auto-riferimento
Das soll	frau	ja tun, besonders in meinem Alter. Bewegung ist wichtig.	Generale + Probabile auto-riferimento
Also praktisch die Heimatuniversität, wo	eine	sich habilitiert hat.	Generale + Probabile auto-riferimento
Und	eine,	die mit einer Frau im Rollstuhl zusammenlebt, die kam damals quasi selbst in den Rollstuhl.	Lesbismo + Probabile auto-riferimento

3.1.2 Sull'utilizzo di sdoppiamento e colpo di glottide

Ci proponiamo di osservare l’uso dello sdoppiamento con colpo di glottide e dello sdoppiamento per esteso all’interno dei due subcorpora, al fine di rilevare se e in che misura i tre parlanti utilizzano queste due strategie di genere.

Colpo di glottide nel parlato di Luise Pusch e degli Intervistatori

Per individuare le occorrenze di colpo di glottide, ci siamo serviti di due query: la prima individua tutte le forme di “innen”, escludendo dai risultati il sostantivo *das Innen* (it: “l’interno”), un possibile falso positivo. La seconda query utilizzata segue lo stesso modello della prima, ma individua le occorrenze di colpo di glottide al singolare. Tutte le occorrenze rilevate sono state ricontrollate tramite il confronto con la traccia audio originale. Nessuna delle due query ha prodotto falsi positivi.

Sdoppiamento per esteso nel subcorpus *Pusch* e nel subcorpus *Intervistatori*

Tra i fattori che complicano la ricerca dello sdoppiamento per esteso figura il gran numero di combinazioni possibili e il fatto che la ricerca di forme che terminano in -in non individua solo sostantivi femminili singolari, ma può rilevare anche elementi irrilevanti per la nostra ricerca, come ad esempio il sostantivo *der Termin* (it: “l’appuntamento”). Per la metodologia di ricerca

dello sdoppiamento all'interno del corpus ci si è ispirati a quella adottata da Elmiger et al. in *Geschlechtergerechte Behördentexte: Linguistische Untersuchungen und Stimmen zur Umsetzung in der mehrsprachigen Schweiz* (2017a), ed è stato utilizzato il seguente modello di query:

[word="*.innen" & tag="N.*Pl.Fem"] [tag="CONJ.*"] []{0,2} [tag="N.*Pl.Masc"]

La query individua i casi in cui un sostantivo femminile plurale (tag="N.*Pl.Fem") che termina in -innen (word="*.innen") è seguito da una qualsivoglia congiunzione [tag="CONJ.*"], che si trova ad una distanza compresa tra 0 e 2 parole ([]{0,2}) da un sostantivo maschile (tag="N.*Pl.Masc").

La scelta di consentire da zero a due elementi aggiuntivi tra la congiunzione e il secondo sostantivo permette di trovare anche forme come “Bürgerinnen und auch viele Bürger” (it: cittadine e anche molti cittadini). Il modello di query è stato leggermente modificato caso per caso, adattandolo alle varie possibili combinazioni (singolare o plurale, maschile-femminile/femminile-maschile, con e senza congiunzione tra i due sostantivi). La rilevanza delle occorrenze è stata poi verificata manualmente.

La tabella 7 riassume i risultati ottenuti tramite l'interrogazione dei subcorpora. La prima colonna indica la strategia di genere utilizzata, la seconda e la terza colonna ne indicano la frequenza assoluta e relativa nel subcorpus *Pusch*, la quarta e la quinta ne indicano la frequenza assoluta e relativa nel subcorpus *Intervistatori*.

Tabella 7 – Sdoppiamento per esteso e colpo di glottide nei due subcorpora

Strategia di genere	Frequenza assoluta in <i>Pusch</i>	Frequenza relativa in <i>Pusch</i> (WpM)	Frequenza assoluta in <i>Intervistatori</i>	Frequenza relativa in <i>Intervistatori</i> (WpM)
Sdoppiamento al singolare	2	75,29	2	130,56
Sdoppiamento al plurale	10	376,48	5	326,39
Colpo di glottide al singolare	0	0,00	0	0,00
Colpo di glottide al plurale	11	414,12	6	391,67
Totale sdoppiamento con colpo di glottide	11	414,12	6	391,67
Totale sdoppiamento per esteso	12	451,77	7	456,95

Confrontando nei due subcorpora l'uso al singolare e al plurale dello sdoppiamento e del colpo di glottide si evince che lo sdoppiamento è stato utilizzato in entrambi i subcorpora prevalentemente al plurale, mentre il colpo di glottide è stato usato in entrambi i subcorpora solo ed esclusivamente al plurale. È possibile osservare che sia intervistata che intervistatori hanno impiegato entrambe le strategie, senza che una prevalesse chiaramente sull'altra, nonostante lo sdoppiamento per esteso sia una strategia consigliata dalla maggior parte delle linee guida e maggiormente affermata in tedesco rispetto al colpo di glottide, che è un fenomeno più recente. In linea di massima, le query individuate si sono rivelate adeguate agli scopi in quanto hanno permesso di individuare le occorrenze di colpo di glottide e di sdoppiamento per esteso, senza produrre falsi positivi.

3.1.3 Sull'uso di aggettivi e participi sostantivati nel corpus *Alles Gesagt?*

La presente sezione è dedicata all'analisi dell'uso di aggettivi e participi sostantivati all'interno del corpus *Alles Gesagt?*.

- Participi presenti

Per la ricerca dei participi presenti ci siamo ispirate alla query utilizzata da Elmiger et al. (2017a:45) a partire dalla desinenza -ende(n), che caratterizza il participio presente in tedesco.

La query elaborata per la ricerca in CQL si basa sul seguente modello:

```
[tag="N.*" & (word=".*ende" | word=".*enden" | word=".*ender")]
```

La query ricerca tutti i sostantivi (tag="N.*")⁶⁸ che terminano in -ende, -enden o -ender. Questa desinenza non è presente solo nei participi presenti, ma anche in una serie di altre parole, come ad esempio *Dividende*, *Gegenden*, *Wochenende* e *Tausende*, il che ne complica la ricerca mirata nel corpus. Nel caso specifico del nostro corpus, il falso positivo più frequente è stato la parola inglese *Gender*, che compare numerose volte per via dell'argomento trattato dall'episodio.

La query è stata dunque migliorata attraverso formule che permettono di evitare la comparsa di determinate parole nei risultati di ricerca⁶⁹. Sono state individuate 3 occorrenze totali di participi

⁶⁸ Elmiger et al. (2017a) ricercano esclusivamente i sostantivi plurali. Nel nostro caso, abbiamo deciso di considerare sia i participi presenti singolari che plurali per ottenere un quadro più completo.

⁶⁹ La sintassi in CQL segue il modello word!="parola da escludere". Per escludere dalla ricerca la parola *Wochenende*, ad esempio, la query andrebbe modificata come segue:

```
[tag="N.*" & (word=".*ende" | word=".*enden" | word=".*ender") & word!="Wochenende"]
```

presenti sostantivati all'interno del corpus. Sono state rilevate 2 occorrenze di participio presente singolare, quindi non inclusivo:

Wer sollte denn jetzt vielleicht mal hier der nächste **Vorsitzende** sein in unserem Verein.

Wenn explizit nach beiden Geschlechtern gefragt wird, wer sollte denn der oder die **Vorsitzende** sein, da kommen immer mehr Frauennamen.

(*Alles Gesagt?*, 22.11.22, 02:03:19 – 02:03:39 Pusch.)

È interessante notare come nel secondo caso, nonostante il participio presente sia al singolare, la neutralità viene garantita attraverso lo sdoppiamento dell'articolo (*der oder die*, sottolineato nell'esempio).

L'unica occorrenza di participio presente sostantivato al plurale è stata individuata nel subcorpus degli intervistatori, e si tratta di un caso "classico", già ampiamente usato in tedesco e consigliato da molte linee guida:

Ich weiß gar nicht wie viele Aufstände es gab, um sie sozusagen in eine Professur zu protestieren von den **Studierenden**.

(*Alles Gesagt?*, 22.11.22, 04:53:00 – 04:53:22 Intervistatori.)

- Participi passati

La ricerca dei participi passati sostantivati è risultata più complessa rispetto ai participi presenti, a causa della varietà di forme differenti in cui il participio passato si presenta in tedesco. In questo caso è stato necessario operare attraverso liste di participi sostantivati. La combinazione di ricerca tramite query e lettura del corpus ha restituito un totale di 6 occorrenze, di cui 3 al singolare (*Gleichstellungsbeauftragte*, *Bekannte*), tutte nel subcorpus dell'intervistata. Sono state individuate 3 occorrenze di participi passati plurali, due nel subcorpus dell'intervistata (*Angestellte*) e una nel subcorpus degli intervistatori (*Frauenbeauftragte*).

La tabella 8 riassume le frequenze relative e assolute dei participi presenti e passati plurali nei due subcorpora.⁷⁰

⁷⁰ Le due occorrenze di participi presenti singolari (75,29 WpM) e le tre occorrenze di participi passati singolari (112,94 WpM), tutte riscontrate nel subcorpus *Pusch*, non sono state considerate nel conteggio.

Tabella 8 – Participi sostantivati nei due subcorpora

Soluzioni di genere	Frequenza assoluta in <i>Pusch</i>	Frequenza relativa in <i>Pusch</i>	Frequenza assoluta in <i>Intervistatori</i>	Frequenza relativa in <i>Intervistatori</i>
Participi presenti plurali	0	0 WpM	1	65,28 WpM
Participi passati plurali	2	75,29 WpM	1	65,28 WpM
Totale participi plurali	2	75,29 WpM	2	130,56 WpM

Analizzando la frequenza di participi presenti e passati sostantivati nei due subcorpora, si nota come sia intervistata che intervistatori abbiano utilizzato entrambi. Osservando però la frequenza d'uso e confrontandola con quella di altre strategie analizzate nelle sezioni precedenti (ad es. colpo di glottide e sdoppiamento), questa strategia sembrerebbe non rappresentare una strategia centrale in nessuno dei due subcorpora.

- Aggettivi sostantivati

La ricerca automatica degli aggettivi sostantivati risulta complicata per via della grande varietà di forme in cui possono presentarsi, e per via del fatto che non sempre queste forme si riferiscono solo a persone (Elmiger et al., 2017a:48). L'identificazione degli aggettivi sostantivati è stata dunque effettuata con l'ausilio di una lista di aggettivi sostantivati comuni in tedesco. Tuttavia, la ricerca non è stata molto produttiva. Sono state individuate solo due occorrenze di aggettivi sostantivati al plurale (*die Jüngeren* e *die Älteren*), entrambe nel corpus dell'intervistata (75,295 WpM).

Anche nel caso degli aggettivi sostantivati, vengono messi in luce i limiti dell'interrogazione di corpora nella ricerca di fenomeni complessi. Oltre alla ricerca tramite liste, un'alternativa potrebbe essere il tagging manuale di aggettivi e participi sostantivati, che ne faciliterebbe la ricerca tramite query. Queste categorie, infatti, spesso vengono taggate semplicemente come sostantivi o come aggettivi. Il tagger GermanRFTagger, ad esempio, li considera come semplici sostantivi e non ne consente la ricerca mirata nel corpus. In un corpus di dimensioni ridotte come il nostro, è possibile leggere il testo per individuare le occorrenze non rilevate tramite la lista di aggettivi sostantivati utilizzata, ma in corpus di dimensioni maggiori questo non è possibile. Osservando i risultati ottenuti, si può affermare che gli aggettivi sostantivati non sembrano rappresentare la strategia di genere principale né per l'intervistata né per gli intervistatori.

3.1.4 Sull'uso di collettivi ed epiceni nel corpus *Alles Gesagt?*

- Collettivi

La ricerca dei collettivi può risultare complicata, oltre che per via della varietà delle forme in cui si presentano, anche perché, per un problema di polisemia, non sempre le voci si riferiscono a persone⁷¹.

Sulla base della metodologia adottata da Elmiger et al. (2017a), è stata effettuata una selezione di voci e ne sono state rilevate le occorrenze nel corpus *Alles Gesagt?*. Le forme con il maggior numero di occorrenze sono riportate nella tabella 10: la colonna di sinistra riporta la lista dei collettivi, la colonna centrale raccoglie le frequenze nel subcorpus *Pusch* e la colonna di destra mostra le frequenze nel subcorpus *Intervistatori*. Sono state prese in considerazione sia forme semplici che composti.

Tabella 9 – Collettivi più frequenti nei due subcorpora

Collettivi (lemmi)	Frequenza assoluta in <i>Pusch</i>	Frequenza relativa in <i>Pusch</i>	Frequenza assoluta in <i>Intervistatori</i>	Frequenza relativa in <i>Intervistatori</i>
<i>Bewegung</i>	26	978,84	10	652,78
<i>Leute</i>	17	640,01	24	1 566,68
<i>Gesellschaft</i>	11	414,13	2	130,56
<i>Partei</i>	9	338,83	1	65,28
<i>Gruppe</i>	7	263,53	2	130,56
<i>Gemeinschaft</i>	6	225,89	0	0,00
<i>Volk</i>	4	150,59	0	0,00
<i>Publikum</i>	2	75,30	1	65,28
<i>Verein</i>	2	75,30	0	0,00
<i>Bevölkerung</i>	2	75,30	0	0,00
<i>Staat</i>	2	75,30	0	0,00
<i>Gremium</i>	2	75,30	0	0,00
<i>Institution</i>	2	75,30	0	0,00
<i>Versammlung</i>	2	75,30	0	0,00
<i>Öffentlichkeit</i>	0	0,00	1	65,28
<i>Menschheit</i>	0	0,00	1	65,28

⁷¹ Ad esempio, *die Opposition* (l'opposizione) può riferirsi a un gruppo di persone, ad esempio quando si parla dei gruppi di opposizione in ambito politico, ma anche avere un significato astratto e non riferito a persone (Elmiger et al.2017:58-59).

Naturalmente, un corpus di dimensioni ridotte come questo non si presta ad un'analisi in larga scala. Dai risultati ottenuti, tuttavia, è comunque possibile trarre alcune osservazioni. Tra i collettivi maggiormente utilizzati nel corpus, molti sono ricollegabili al campo semantico della lotta politica femminista (*Bewegung, Gesellschaft, Gruppe, Gemeinschaft, Partei*). Il termine *Bewegung* compare ad esempio non solo nella sua forma semplice, ma anche in forme composte come *Frauenbewegung, Lesben- und Schwulenbewegung, Queer-Bewegung, Now-Bewegung, Männerbewegung, Erweckungsbewegung, Sex-Positiv-Bewegung, Bürgerrechtsbewegung e Fahnenbewegung*. Analogamente, il sostantivo *Gemeinschaft* compare non solo in forma semplice ma anche in composti come *Frauengemeinschaft, Männergemeinschaft e Sprachgemeinschaft*, mentre il sostantivo *Gruppe* compare anche come composto in *Frauengruppe*.

Un altro collettivo molto frequente è *Leute*, che può essere usato per riferirsi a qualsiasi gruppo di persone, alla pari di *die Personen* o *die Menschen*,. Tra i collettivi più utilizzati, diversi presentano delle desinenze tipicamente associate ai sostantivi collettivi, come *-schaft* (*Gemeinschaft, Gesellschaft*), *-ung* (*Bewegung, Bevölkerung, Versammlung*), o anche *-heit* (*Menschheit*) e *-keit* (*Öffentlichkeit*). Altrettanto numerosi sono però i collettivi che non seguono questo modello (*Volk, Leute, Gruppe* etc.). L'analisi dei collettivi, per quanto presenti nei limiti, ha permesso di evidenziare che entrambe le parti coinvolte nell'interazione hanno utilizzato questa strategia di genere, seppur con modalità diverse.

- Epiceni

Analogamente a quanto detto per i collettivi, anche i sostantivi epiceni sono caratterizzati dalla grande varietà delle forme in cui si presentano, e non sempre si riferiscono a persone (ad esempio *Individuum* può anche riferirsi ad animali e piante, nell'ambito della biologia).

Per interrogare i due subcorpora è stata effettuata una selezione di sostantivi epiceni comuni in tedesco. Non è stato rilevato un alto numero di occorrenze, fatta eccezione per i tre epiceni *Kind, Mensch* e *Person*. La tabella 11 raccoglie le occorrenze degli epiceni con le frequenze maggiori tra quelli ricercati all'interno del corpus.

Tabella 10 – Epiceni maggiormente frequenti nei due subcorpora

Epiceni (lemmi)	Frequenza assoluta in <i>Pusch</i>	Frequenza relativa in <i>Pusch</i>	Frequenza assoluta in <i>Intervistatori</i>	Frequenza relativa in <i>Intervistatori</i>
<i>Kind</i>	34	1 280,02	21	1 370,79
<i>Mensch</i>	10	376,48	15	979,96
<i>Person</i>	6	225,89	6	391,67
<i>Opfer</i>	6	225,89	1	65,28
<i>Figur</i>	2	75,30	0	0,00
<i>Genie</i>	1	37,65	2	130,56
<i>Individuum</i>	1	37,65	0	0,00
<i>Star</i>	0	0,00	1	65,28
<i>Fan</i>	0	0,00	1	65,28

Confrontando le frequenze di *Mensch*, *Person* e *Individuum*⁷², *Mensch* sembra essere il più utilizzato tra i tre epiceni, e compare anche in composti come *Audiomenschen* e *CDU-Mensch*. L'osservazione delle occorrenze ha rivelato che *Mensch* compare soprattutto al plurale in entrambi i subcorpora. *Person* compare invece prevalentemente al singolare. Tra gli epiceni rilevati troviamo anche il sostantivo *Kind*, che presenta il maggior numero di occorrenze in assoluto.

Combinazione di epiceni e aggettivi

Osservando i contesti d'uso degli epiceni presi in esame, si nota come alcuni epiceni vengano combinati con aggettivi o participi, sia dall'intervistata che dagli intervistatori. Nel corpus dell'intervistata troviamo forme come *verstorbene Personen* e *göttliche Figuren* (invece di *die Verstorbenen* e *die Götter/Göttinnen*), mentre gli intervistatori parlano ad esempio di *akademisch gebildete Menschen* (invece di *die Akademiker/innen*) e usano anche riformulazioni del tipo “Menschen, [...] die *gestorben* sind” (invece di *die Verstorbenen*). Inoltre, nel corpus degli intervistatori troviamo anche alcune soluzioni al singolare, come *die dritte Person* (invece di *der/die Dritte*) e *die einzige Person* (invece di *der/die Einzige*).

⁷² Sul modello della metodologia impiegata da Elmiger et al. (2017a:49-58), che confronta le frequenze dei tre termini in quanto vengono considerati come sinonimi.

3.1.5 Alcuni pronomi indefiniti e la loro ripresa anaforica

Una soluzione proposta da diverse linee guida consiste nell'uso di pronomi come *alle*, *jemand*, *wer*, *diejenige*. Tuttavia, in alcuni casi, quando pronomi grammaticalmente maschili come *jemand*, *niemand* e *wer* vengono ripresi, viene esplicitato il genere maschile (ad es. *Niemand kann **sein** Schicksal beeinflussen*). In tal caso, il pronome talvolta non viene più percepito come neutrale, ma come maschile (Cancelleria federale svizzera, 2023:10). La linguistica femminista già negli anni Ottanta proponeva la ripresa di pronomi come *wer* anche al femminile (Pusch, 1984:151-153). Uno studio condotto da Ellsäßer nel 2023 affronta il tema della ripresa al femminile di *jemand* (ad es. „Sie ist jemand, die gerne Kaffee trinkt“) dimostrando che il fenomeno è relativamente raro, ma comunque presente. L'autrice argomenta che l'accordo al femminile può essere ricollegato a motivazioni pragmatiche, cioè legate alle intenzioni comunicative del/la parlante, ad esempio nel caso in cui il/la parlante cerca di utilizzare un linguaggio sensibile al genere.

Osservando le riprese di questi pronomi all'interno dei due subcorpora, è stato individuato un singolo caso di ripresa al femminile, relativo al pronome *jemand*:

Also ich glaube, wenn man denkt, also Schreiben ist nicht so direkt mein Metier, dann ist es günstig, wenn man **jemanden** hat, die dann immer Hurra ruft oder sagt, ja, weiter so, das ist, ja, förderlich.

(*Alles Gesagt?*, 22.11.22, 03:13:05 – 03:13:20, Pusch.)

In generale, è risultato più frequente l'uso dei pronomi senza alcuna ripresa, e qualora questi sono stati ripresi, è prevalsa la ripresa al maschile.

Riallacciandoci all'interpretazione pragmatica di Ellsäßer (2023) di questo fenomeno, è possibile pensare che anche l'unica occorrenza dell'accordo al femminile, individuata nel subcorpus *Pusch*, possa ricollegarsi alla volontà di utilizzare un linguaggio inclusivo, soprattutto alla luce del background della linguista.

3.1.6 Le strategie di genere utilizzate a confronto

Riprendendo le strategie di genere elencate nelle tabelle 1 e 2 al capitolo 1, si riassumono nella colonna “commenti” della Tabella 13 le principali riflessioni tratte dalle analisi portate avanti sul corpus *Alles Gesagt?*.

Tabella 11 – Le strategie di genere analizzate nel corpus *Alles Gesagt*?⁷³

Strategia di genere	Commenti
Aggettivi sostantivati al plurale	Gli aggettivi sostantivati sono stati utilizzati prevalentemente dall'intervistata e non sembrano costituire una strategia centrale di linguaggio inclusivo nel parlato analizzato.
Participi sostantivati al plurale	Sono stati utilizzati sia dall'intervistata che dagli intervistatori, ma non sembrano costituire una strategia di genere centrale nel parlato analizzato. Gli intervistatori hanno utilizzato esclusivamente forme al plurale, mentre l'intervistata ricorre anche a forme singolari. Tuttavia, l'uso dello sdoppiamento dell'articolo (come in <i>der oder die Vorsitzende</i>) da parte dell'intervistata mostra che anche i participi singolari possono essere impiegati in modo inclusivo. Nel complesso, le forme riscontrate sono perlopiù soluzioni già consolidate e consigliate dalle linee guida.
Sostantivi epiceni e suffissi neutri	Sia intervistata che intervistatori hanno utilizzato sostantivi epiceni. Tra questi, oltre a sostantivi tipicamente consigliati dalle linee guida come <i>Person</i> e <i>Mensch</i> , troviamo anche il sostantivo epiceno <i>Kind</i> , che risulta la forma più usata in assoluto.
Sostantivi collettivi	Tutti i partecipanti all'interazione hanno utilizzato sostantivi collettivi, seppur con modalità diverse. Tra questi spiccano per frequenza in entrambi i subcorpora i sostantivi <i>Leute</i> e <i>Bewegung</i> .
Sdoppiamento per esteso e colpo di glottide⁷⁴	Entrambe le strategie sono presenti nel parlato sia dell'intervistata sia degli intervistatori. Nessuna delle due strategie prevale in modo netto sull'altra, nonostante lo sdoppiamento per esteso sia generalmente più consolidato e raccomandato dalle linee guida. In entrambi i subcorpora, lo sdoppiamento per esteso è impiegato soprattutto al plurale, mentre il colpo di glottide compare esclusivamente al plurale.
Pronomi indefiniti	Alcuni pronomi indefiniti, come <i>wer</i> , <i>jemand</i> e <i>niemand</i> , sono grammaticalmente maschili e prevedono nella ripresa anaforica l'accordo al maschile. La possibilità di ripresa al femminile, per quanto propugnata dalla stessa Luise Pusch già negli anni Ottanta, rimane un fenomeno molto marginale, come già osservato da Ellsäßer (2023).

⁷³ Alcune soluzioni elencate nelle Tabelle 1 e 2, come il ricorso alla diatesi passiva o a riformulazioni impersonali, e l'uso della forma di cortesia non sono state analizzate.

⁷⁴ Nella tabella iniziale erano presenti le strategie di *splitting*, l'uso della *Binnen-I* e dei segni grafici tipici dello scritto, mentre la nostra analisi si è concentrata sulla loro controparte orale, il colpo di glottide.

Evitare l'uso di <i>man</i>	L'analisi dell'uso dei pronomi indefiniti <i>frau</i> e <i>man</i> nel corpus <i>Alles Gesagt?</i> ha mostrato una netta asimmetria tra il parlato dell'intervistata Luise Pusch e quello degli intervistatori. La linguista ha utilizzato <i>frau</i>, <i>eine</i> e <i>keine</i>, ma <i>man</i> rimane comunque largamente predominante, con un rapporto complessivo di 1:10 a favore di <i>man</i>. Gli intervistatori, invece, fanno ricorso esclusivamente a <i>man</i> e non utilizzano <i>frau</i>. Di conseguenza, è possibile affermare che gli intervistatori <u>non</u> hanno utilizzato affatto questa strategia di genere. L'osservazione qualitativa dei contesti d'uso rivela che i pronomi <i>frau</i> e <i>eine</i> sono stati utilizzati dalla linguista prevalentemente in contesti marcati dal punto di vista di genere, come il femminismo e il lesbismo, e/o in situazioni di auto-riferimento.
Aggettivi con sostantivi epiceni	Sia intervistata che intervistati hanno talvolta combinato l'uso di epiceni con aggettivi, come nel caso di <i>verstorbene Personen</i>.

3.2 Analisi del corpus Politische Reden (1982-2020)

3.2.1 Alternative a *man*

Contrariamente a quanto riscontrato nel corpus *Alles Gesagt?*, la ricerca nel corpus *Politische Reden (1982-2020)* non ha rivelato alcuna occorrenza del pronome *frau*⁷⁵, mentre sono state individuate 65505 occorrenze (2426,11 WpM) del pronome *man*.⁷⁶

3.2.2 Sdoppiamento in forma contratta

La ricerca tramite query⁷⁷ ha permesso di individuare le occorrenze di *splitting*, segni grafici e *Binnen-I* nel corpus *Politische Reden (1982-2020)*, confrontabili nella figura 2. Lo sdoppiamento con *Binnen-I* (305 occorrenze totali, 11,30 WpM) risulta essere di gran lunga la modalità più utilizzata tra quelle prese in considerazione. Tra i lemmi più frequentemente sdoppiati con *Binnen-I* figurano *Arbeitnehmerin* (25 occorrenze, 0,93 WpM), *Kollegin* (20 occorrenze, 0,74 WpM), *Mitarbeiterin* (15 occorrenze, 0,56 WpM), *Alleinerzieherin* (12 occorrenze, 0,44 WpM) e *Lehrerin* (11 occorrenze, 0,41 WpM).

La figura 3 permette di osservare le frequenze assolute relativamente all'uso della *Binnen-I*, dello *splitting* e del trattino basso nel corso degli anni. Notiamo come la *Binnen-I* compare nel corpus solo a partire dal 1997 e resta molto sporadica, pur con fluttuazioni annuali, per crescere a partire dagli anni 2000 fino a raggiungere un picco registrato nel 2018 e poi calare di nuovo. Lo *splitting* compare con maggiore costanza dal 2005 in poi, senza però mai raggiungere le frequenze d'uso della *Binnen-I*. Il trattino basso conta solo 4 occorrenze totali, di cui 3 nel 2018 e una nel 2014.

I dati attuali non permettono di conoscere i motivi del calo improvviso tra il 2018 e il 2020; tra le ipotesi da verificare figurano il passaggio a nuove convenzioni di trascrizione o a un maggiore uso dello sdoppiamento per esteso. Anche l'introduzione in Germania del riconoscimento del genere "divers" nel 2018 potrebbe aver incentivato un uso maggiore di

⁷⁵ Tuttavia, la ricerca del pronome *frau* con lettera maiuscola non è stata effettuata per via delle problematiche di disambiguazione e di efficacia della query illustrate nell'analisi dello stesso fenomeno nel corpus *Alles Gesagt?*, per cui non è possibile escludere l'eventuale presenza di occorrenze di *frau* a inizio frase.

⁷⁶ La ricerca dei pronomi *einer*, *eine*, *keiner* e *keine* non è stata effettuata.

⁷⁷ Per la ricerca della *Binnen-I* al plurale e al singolare sono state usate le query ***Innen WITHOUT Innen** e ***In WITHOUT In**, che permettono di individuare le parole che terminano in *-Innen*, o in *-In*, escludendo i falsi positivi del sostantivo *Innen* (it: "l'interno") e della preposizione *In* (it: "dentro"). Per i segni grafici e lo slash è stato inserito un backslash tra l'asterisco all'inizio della query e il simbolo da ricercare (ad esempio, per trovare casi di *Gendersternchen* al plurale, la query è ****innen**). Il backslash (\) permette di cercare slash e segni grafici, che altrimenti vengono interpretati dal sistema come operatori con funzioni specifiche.

convenzioni che includono anche le persone non binarie e un conseguente calo nell'uso della *Binnen-I*.

Figura 2 – Tipi di sdoppiamento in forma contratta nel corpus

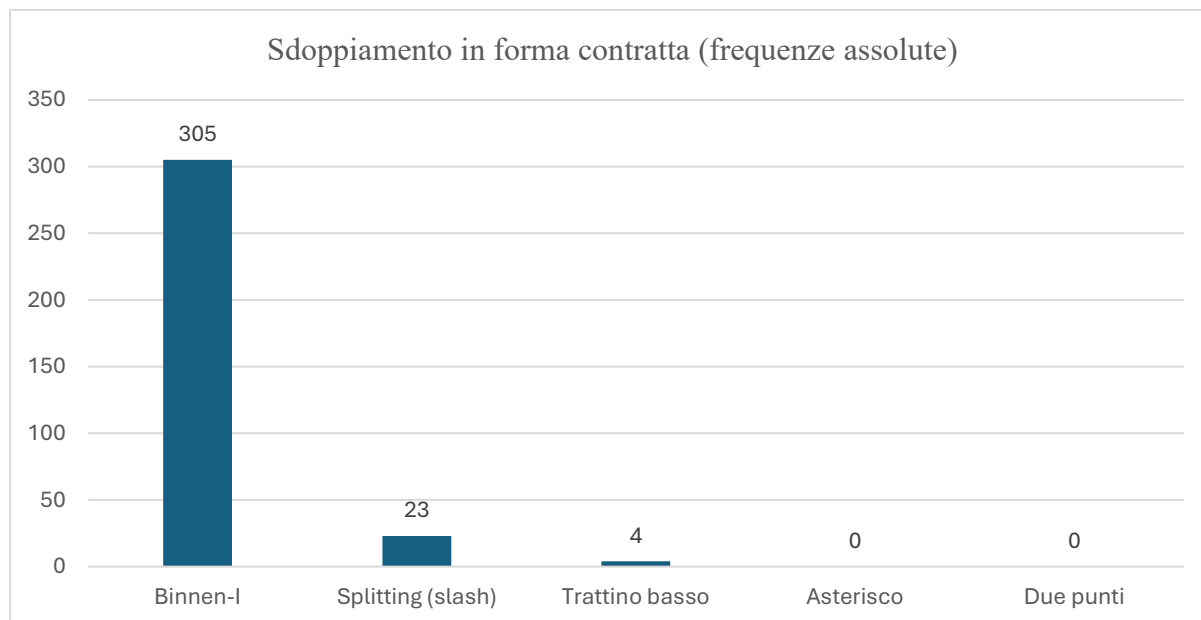
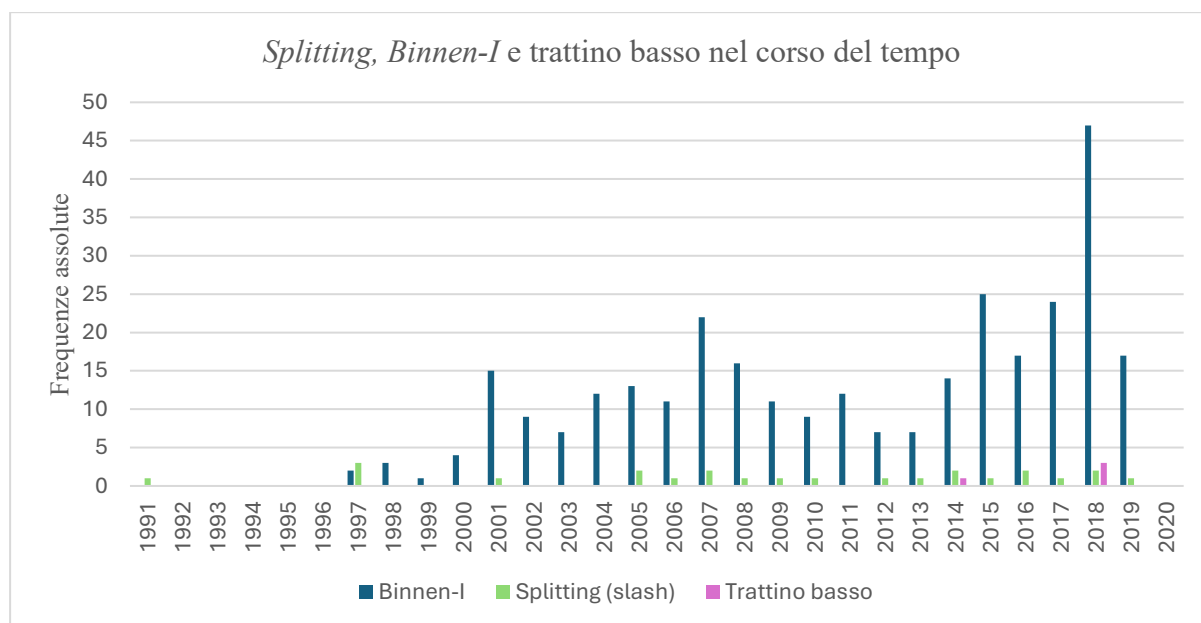


Figura 3 - Uso di *splitting*, trattino basso e *Binnen-I* all'interno del corpus nel corso del tempo



3.2.3 Analisi per coppie di *Personenbezeichnungen*

Oltre che per via della grande varietà di possibili combinazioni, la ricerca dello sdoppiamento per esteso all'interno del corpus *Politische Reden (1982-2020)* risulta più complessa rispetto al corpus *Alles Gesagt?* in quanto per l'analisi dei corpora del DWDS non sembrano essere disponibili tag appositi per il genere o il numero dei sostantivi. La ricerca è stata ridimensionata a tre coppie di *Personenbezeichnungen* selezionate tra quelle più frequenti all'interno del corpus *Politische Reden (1982-2020)* e riportate di seguito:

Elemento 1	Elemento 2
<i>Bürgerin</i>	<i>Bürger</i>
<i>Mitarbeiterin</i>	<i>Mitarbeiter</i>
<i>Wählerin</i>	<i>Wähler</i>

La disambiguazione tra l'uso del maschile generico e l'uso specifico spesso non è possibile dal momento che richiede informazioni sul contesto situativo non sempre disponibili, pertanto ci limitiamo al confronto tra le occorrenze delle sole forme maschili e quelle di entrambe le forme maschile e femminile. Oltre al confronto tra forme sdoppiate e maschili, vengono anche prese in considerazione altre soluzioni di genere, laddove applicabili, come participi presenti e passati sostantivati e sostantivi collettivi.

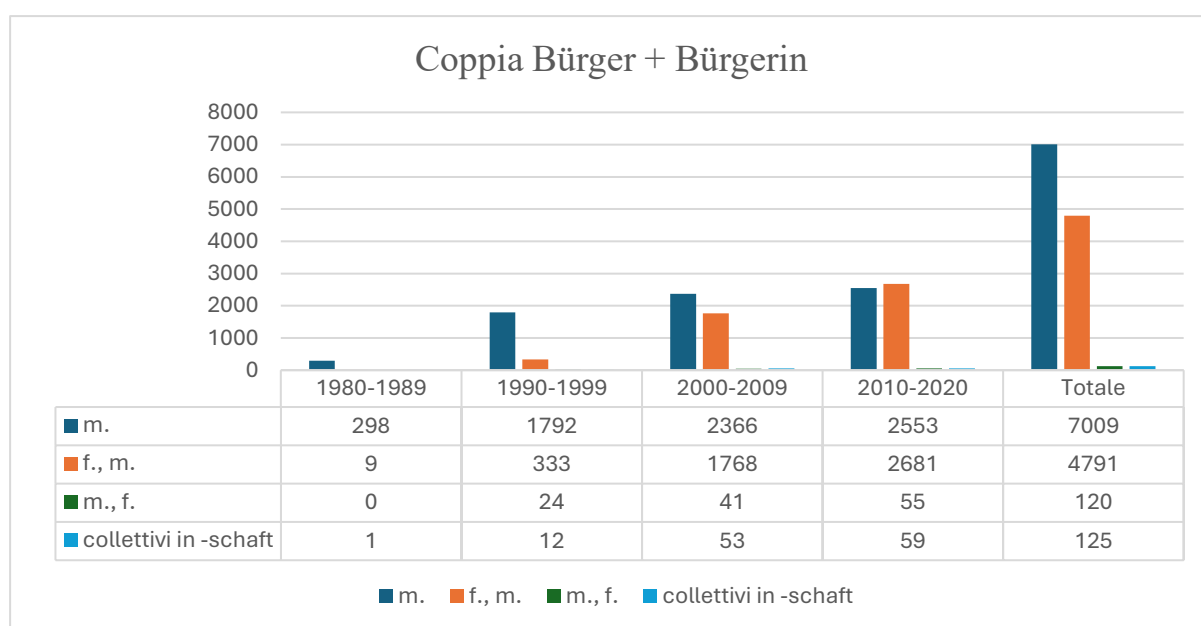
Le occorrenze delle forme più frequenti per ogni coppia vengono analizzate sia confrontandone la frequenza totale sia in prospettiva diacronica, al fine di evidenziare tendenze e sviluppi nel loro uso all'interno del corpus. Per risultati inferiori alle 500 occorrenze, tutte le occorrenze sono state tutte verificate manualmente; per risultati superiori alle 500 occorrenze si è operato tramite un controllo a campione tra le 300 e le 500 occorrenze.

Coppia *Bürgerin* + *Bürger*

Per la coppia *Bürgerin* + *Bürger* sono state analizzate la forma al maschile (*m.*), lo sdoppiamento nell'ordine maschile-femminile (*m., f.*), lo sdoppiamento nell'ordine femminile-maschile (*f., m.*) e la forma collettiva *Bürgerschaft* (collettivi in *-schaft*).

La figura 3 mostra che il maschile prevale sulle altre forme in termini assoluti (totale delle occorrenze). Analizzando l'uso delle varie forme nel tempo, si nota una netta predominanza iniziale della forma maschile (*m.*). A partire dagli anni Novanta aumenta però in modo significativo anche l'uso delle forme inclusive sdoppiate, in particolare nell'ordine femminile-maschile (*f., m.*). È interessante notare come nel periodo 2010-2020 la forma femminile (2681, 99.29 WpM) abbia raggiunto e persino superato di poco quella maschile (2553, 94.55 WpM), segnando un cambiamento significativo nel corso degli anni. La forma sdoppiata nell'ordine maschile-femminile (*m., f.*) e la forma collettiva in *-schaft* rimangono invece comparativamente marginali lungo tutto il periodo.

Figura 4 – Forma maschile e forme alternative in ottica diacronica per la coppia *Bürger* + *Bürgerin* (frequenze assolute)

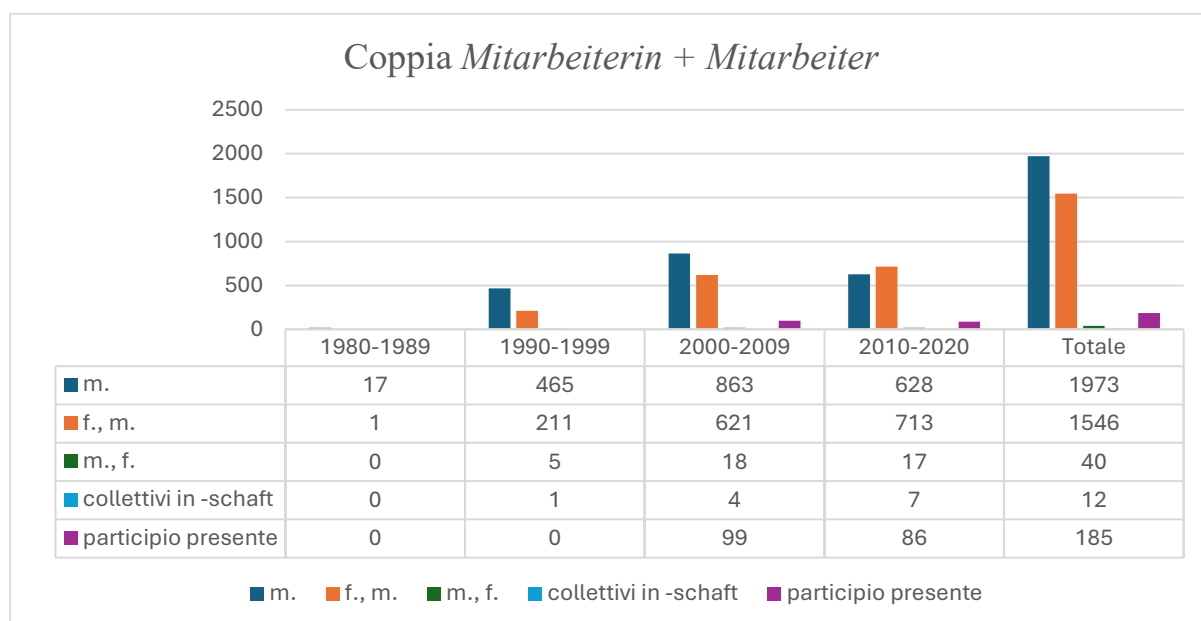


Coppia *Mitarbeiterin* + *Mitarbeiter*

Per la coppia *Mitarbeiterin* + *Mitarbeiter* sono state analizzate la forma al maschile (*m.*), lo sdoppiamento nell'ordine maschile-femminile (*m., f.*), lo sdoppiamento nell'ordine femminile-maschile (*f., m.*) la forma collettiva *Mitarbeiterschaft* e la forma al participio presente *Mitarbeitende*.

Anche in questo caso, dai dati riportati in figura 4 si evince che il maschile prevale sulle altre forme per quanto riguarda il numero di occorrenze totali. Tuttavia, osservando lo sviluppo dei dati nel tempo si osserva l'introduzione progressiva di nuove forme inclusive: mentre il maschile dominava fino al 2009, il decennio 2010-2020 mostra un'inversione di tendenza a favore dello sdoppiamento *f., m.*, il cui uso cresce a partire dagli anni Novanta fino a superare per frequenza il maschile (713 occorrenze, 26,41 WpM contro 628 occorrenze, 23,26 WpM). Inoltre, a partire dagli anni Duemila compare la forma *Mitarbeitende* (participio presente) come soluzione di neutralizzazione, che mantiene una presenza costante (intorno alle 100 unità, circa 3,70 WpM) tra il 2000 e il 2020, mentre la forma collettiva *Mitarbeiterschaft* e lo sdoppiamento mf. rimangono fenomeni marginali.

Figura 5 – Forma maschile e forme alternative in ottica diacronica per la coppia *Mitarbeiterin* + *Mitarbeiter* (frequenze assolute)

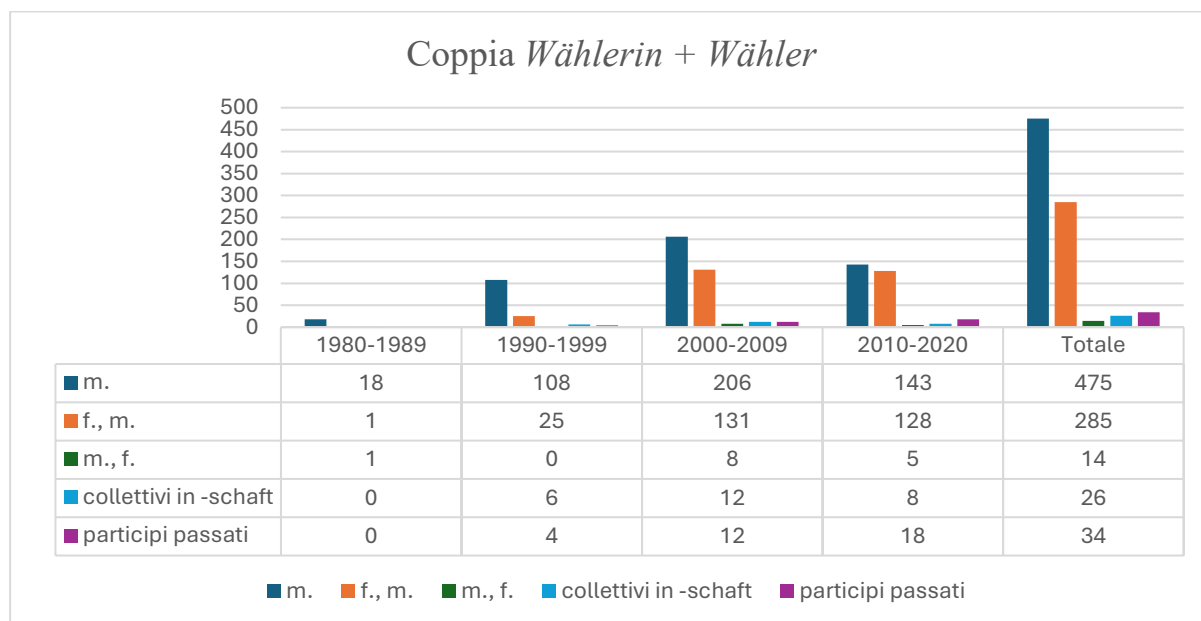


Coppia *Wählerin* + *Wähler*

Per la coppia *Wählerin* + *Wähler* sono state analizzate la forma al maschile (*m.*), lo sdoppiamento nell'ordine maschile-femminile (*m., f.*)⁷⁸ e femminile-maschile (*f., m.*), la forma collettiva *Wählerschaft* e la forma al participio passato *Wahlberechtigte*⁷⁹.

Nella figura 5 si osserva come il maschile domini sulle altre forme per quanto riguarda le occorrenze totali. A livello diacronico, la forma maschile è quasi l'unica rilevata negli anni Ottanta e il numero delle occorrenze documentate cresce progressivamente fino a raggiungere il suo picco massimo nel decennio 2000-2009. Parallelamente, però, cresce anche l'utilizzo della forma sdoppiata *f.,m.* tra gli anni Novanta e l'inizio degli anni Duemila. Nell'ultimo periodo (2010-2020), si nota un calo della forma maschile a favore di una maggiore parità numerica con la forma sdoppiata nell'ordine femminile-maschile (143 occorrenze, 5,29 WpM contro 128 occorrenze 4,74 WpM), indicando una stabilizzazione verso un linguaggio più bilanciato rispetto agli anni Ottanta e Novanta, mentre la forma sdoppiata nell'ordine maschile-femminile e la forma collettiva in *-schaft* rimangono marginali. Un altro dato interessante è la comparsa, a partire dagli anni Novanta, del participio passato sostantivato *Wahlberechtigte*, che mostra una crescita d'uso tra gli anni Novanta e il 2020.

Figura 6 – Forma maschile e forme alternative in ottica diacronica per la coppia *Wählerin* + *Wähler* (frequenze assolute)



⁷⁸ Tutte le occorrenze sono state verificate manualmente e non hanno riportato falsi positivi.

⁷⁹ *Wählerschaft* e *Wahlberechtigte* sono sinonimi di *Wähler* secondo il DWDS (<https://www.dwds.de/wb/W%C3%A4hler>).

Conclusioni sull'analisi del corpus *Politische Reden (1982-2020)*

L'assenza del pronome *frau* nel corpus *Politische Reden (1982-2020)* e la sua presenza nel corpus *Alles Gesagt?* permettono di riflettere su come, nella ricerca di fenomeni linguistico-comunicativi, sia di fondamentale importanza non solo la composizione quantitativa ma anche quella qualitativa. Confrontando le occorrenze di sdoppiamento in forma contratta su tutto il corpus con le occorrenze di sdoppiamento per esteso prese in analisi risulta chiaro che lo sdoppiamento in forma contratta non rappresenta la strategia di genere prevalente. L'analisi in chiave diacronica sia dell'uso della *Binnen-I* che delle varie soluzioni di genere per le tre coppie prese in esame permette di gettare luce sugli sviluppi del linguaggio di genere nei discorsi politici raccolti nel corpus *Politische Reden (1982-2020)*, mostrando la comparsa di alternative al maschile generico e la loro evoluzione nel tempo. In particolare, lo sdoppiamento nell'ordine femminile maschile risulta essere la strategia inclusiva maggiormente impiegata nei tre casi analizzati, talvolta raggiungendo o addirittura superando l'uso del maschile generico nel corpus, mentre altre forme, seppur presenti, non raggiungono frequenze elevate. I limiti dell'analisi, tra cui la necessità di limitarsi a tre coppie di *Personenbezeichnungen* e l'impossibilità di verificare manualmente tutte le occorrenze per risultati superiori alle 500 occorrenze suggeriscono cautela nel trarre conclusioni di portata generale. Ad ogni modo, l'interrogazione del corpus *Politische Reden (1982-2020)* ha permesso di osservare l'evoluzione del linguaggio di genere in un contesto comunicativo differente da quello dell'intervista del corpus *Alles Gesagt?*, su una quantità di dati maggiore e in prospettiva diacronica.

3.3 Discussione dei risultati

I risultati delle due analisi condotte aprono la strada ad alcune riflessioni sull'uso del linguaggio di genere nel tedesco parlato in due corpora che si distinguono per dimensioni, ambito, registro e grado di formalità, genere e situazione comunicativa, livello presumibile di pianificazione linguistica, arco temporale di riferimento, modalità di parlato (dialogico/monologico) e numero di parlanti.

Nel corpus *Alles Gesagt?* emerge una notevole varietà di strategie di genere. L'analisi qualitativa mostra come l'intervistata e gli intervistatori abbiano fatto ricorso talvolta alle stesse soluzioni, talvolta a soluzioni differenti. L'intervistata, che dimostra una forte sensibilità verso il linguaggio di genere, utilizza anche soluzioni più "radicali", come il ricorso ai pronomi *frau* e *eine* al posto di *man* e *einer*, e la ripresa al femminile di *jemand*, mentre gli intervistatori

utilizzano soluzioni di genere più ancorate all'uso linguistico dominante e a quanto generalmente suggerito dalle linee guida.

Sia intervistatori che intervistata hanno utilizzato ampiamente lo sdoppiamento per esteso e il colpo di glottide, i sostantivi epiceni e i collettivi, mentre i participi e gli aggettivi sostantivati non sembrano rappresentare una strategia centrale nel parlato analizzato. Ad ogni modo, nessuna strategia risulta nettamente dominante sulle altre: al contrario, il linguaggio di genere appare come un repertorio flessibile di possibilità, attivate in funzione della situazione comunicativa e delle intenzioni pragmatiche dei parlanti.

Un dato particolarmente significativo, la cui analisi è stata resa possibile grazie all'ascolto della traccia originale, è l'uso del colpo di glottide, un fenomeno al momento ancora non ampiamente studiato, che nel corpus compare con una frequenza paragonabile a quella dello sdoppiamento per esteso, nonostante si tratti di una strategia relativamente recente e meno istituzionalizzata rispetto ad altre.

Il corpus *Politische Reden (1982-2020)* offre una prospettiva diacronica che si configura come complementare all'analisi del primo corpus, consentendo di osservare l'uso del linguaggio di genere nel parlato politico di area germanofona lungo un arco temporale di quasi quarant'anni e offrendo la possibilità di analizzare in che misura il linguaggio politico tedesco riflette i cambiamenti sociolinguistici legati al linguaggio di genere che si sono susseguiti nel tempo.

L'assenza del pronome *frau* nel corpus *Politische Reden (1982-2020)* e la sua presenza nel corpus *Alles Gesagt?* permettono di riflettere su come, nella ricerca di fenomeni linguistico-comunicativi, sia di fondamentale importanza non solo la composizione quantitativa ma anche quella qualitativa.

L'analisi delle varie forme di sdoppiamento evidenzia una netta predominanza nel corpus *Politische Reden (1982-2020)* dello sdoppiamento per esteso sullo sdoppiamento in forma contratta e, tra le forme contratte, una prevalenza della *Binnen-I*. Quest'ultima forma sembra crescere fino al 2018, per poi subire un calo improvviso; tuttavia, per leggere correttamente questi dati e individuare delle tendenze occorrerebbe uno studio molto più approfondito.

L'analisi diacronica dei dati relativi alle tre coppie di *Personenbezeichnungen* selezionate evidenzia una progressiva diffusione, soprattutto a partire dagli anni Novanta, di forme alternative al maschile generico, in particolare lo sdoppiamento nell'ordine femminile–maschile, il cui uso aumenta nel tempo fino a raggiungere (e in alcuni casi superare) le

frequenze del maschile generico, mentre strategie neutralizzanti come i participi sostantivati e i sostantivi collettivi compaiono in misura minore.

Sebbene si tratti di un corpus di lingua parlata, il corpus *Politische Reden (1982-2020)* non offre la possibilità diretta di ascoltare le tracce audio originali, e quindi non permette l'analisi immediata di fenomeni puramente orali come l'uso del colpo di glottide, il che conferma l'importanza, nell'analisi del parlato, dell'accesso diretto alle registrazioni audio/video originali.

I limiti dell'analisi, tra cui la dimensione ridotta del corpus *Alles Gesagt?*, la selezione limitata di *Personenbezeichnungen* analizzati in chiave diacronica e l'impossibilità di verificare manualmente tutte le occorrenze per totali superiori alle 500 occorrenze suggeriscono cautela nel trarre conclusioni di portata generale.

Ciononostante, la combinazione di analisi qualitativa e quantitativa e l'interrogazione di corpora differenti si è rivelata particolarmente fruttuosa, in quanto ha permesso di esplorare gli stessi fenomeni da punti di vista differenti e di lavorare su quantità di dati maggiori. È stato possibile non solo evidenziare delle tendenze d'uso, ma anche riflettere sui contesti d'uso di determinate scelte linguistiche, nonché sul rapporto tra consapevolezza di genere e uso effettivo del linguaggio di genere. Nel complesso, l'analisi ha confermato che il linguaggio di genere non si configura come un sistema unitario e coerente, bensì come un insieme dinamico di pratiche linguistiche la cui realizzazione dipende dal contesto e dalle scelte dei/delle partecipanti all'interazione.

Conclusioni

Il presente elaborato si è posto l'obiettivo di analizzare, alla luce della crescente importanza dell'uso del linguaggio di genere nella comunicazione, l'impiego di alcune strategie di genere all'interno di due corpora di lingua tedesca parlata. L'analisi del corpus *Alles Gesagt?* ha permesso di osservare l'uso in un contesto dialogico informale di alcune delle principali soluzioni consigliate dalle linee guida per il tedesco, sottolineando la rilevanza del background di chi parla nella scelta delle soluzioni di genere da applicare, nonché gettando luce sull'ampio ventaglio di soluzioni possibili che vengono attivate dai/lle parlanti in funzione delle diverse circostanze comunicative, invece di ricorrere ad una singola soluzione come una sorta di panacea.

L'analisi in chiave diacronica del corpus *Politische Reden (1982-2020)* ha permesso di tracciare l'evoluzione dell'uso di determinate soluzioni inclusive nell'ambito dei discorsi di personalità politiche di spicco di paesi di lingua tedesca nel corso degli ultimi decenni, evidenziando, sulla base dell'analisi di tre coppie di sostantivi, come alcune forme inclusive, praticamente assenti negli anni Ottanta del secolo scorso, siano state protagoniste di una crescita graduale nell'uso tra gli anni Novanta e il 2020 fino ad arrivare persino a superare nell'uso la forma maschile.

Lo studio fornisce anche spunti d'interesse per possibili studi futuri.

Dal punto di vista metodologico, la ricerca ha evidenziato l'efficacia delle tecnologie di intelligenza artificiale nel facilitare la trascrizione e l'analisi del parlato, offrendo una metodologia replicabile per futuri studi sull'oralità. Inoltre, la combinazione di corpora differenti ha offerto un quadro d'analisi più completo e ha permesso l'osservazione dello stesso fenomeno su una base di dati maggiore e sotto diversi punti di vista.

Questa metodologia, qui applicata a poche situazioni comunicative e un ridotto numero di strutture linguistiche, potrebbe essere estesa all'osservazione di altri contesti comunicativi. In particolare, lo studio condotto sul corpus *Alles Gesagt?* si configura come uno studio pilota, in piccola scala, ma potrebbe essere interessante confrontare le strategie di genere impiegate da diversi/e intervistati/e all'interno del podcast; analogamente, l'analisi del corpus *Politische Reden (1982-2020)* potrebbe essere arricchita da un confronto delle strategie prevalenti nei testi di parlanti provenienti dall'uno o dall'altro paese, o tra le strategie privilegiate da figure politiche afferenti all'uno o all'altro partito. Resta di grande interesse anche osservare l'evoluzione delle forme inclusive binarie e non binarie nel periodo post-2020. Nell'ottica della

dicotomia scritto-orale, potrebbe essere d'interesse anche estendere l'analisi a corpora di lingua scritta e realizzare un confronto di tendenze d'uso tra la lingua scritta e la lingua orale.

I risultati ottenuti possono essere d'interesse per la formazione di interpreti e traduttori, in quanto evidenziano la diffusione e la rilevanza, perlomeno nei contesti presi in esame, delle forme inclusive, suggerendo la necessità di implementare “automatismi di genere” nella pratica interpretativa, al fine di garantire una resa che sia al contempo rispettosa e fruibile.

In questo senso è auspicabile un ampliamento della ricerca sull'uso del colpo di glottide, fenomeno ancora limitatamente studiato, che in alcuni contesti comunicativi viene ampiamente impiegato e che, grazie alla sua intrinseca economicità, combinato con altre forme maggiormente diffuse e standardizzate, potrebbe rappresentare una strategia preziosa anche per il parlato degli/delle interpreti.

Bibliografia

- Ayaß, R. (2008). *Kommunikation und Geschlecht: Eine Einführung*. W. Kohlhammer.
- Baker, P., & McEnery, T. (a cura di). (2015). *Corpora and discourse studies: Integrating discourse and corpora*. Palgrave Macmillan. <https://doi.org/10.1057/9781137431738>.
- Barbaresi, A. (2018). A corpus of German political speeches from the 21st century. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 792–797). European Language Resources Association (ELRA).
- Baroni, M. (2010). Corpora di italiano. In *Enciclopedia dell'Italiano*. Istituto della Enciclopedia Italiana Treccani.
- Bast, J., Maier, J., Albert, G., & Schneider, J. (2024). Gendered debates? The use of gender-sensitive language in German televised debates, 1997–2022. *European Journal of Politics and Gender*. <https://doi.org/10.1332/25151088y2024d000000023>.
- Bazzanella, C. (2010). Genere e lingua. In *Enciclopedia dell'Italiano*. Istituto della Enciclopedia Italiana Treccani. [https://www.treccani.it/enciclopedia/genere-e-lingua_\(Enciclopedia-dell%27Italiano\)/](https://www.treccani.it/enciclopedia/genere-e-lingua_(Enciclopedia-dell%27Italiano)/).
- Behaghel, O. (1927). Geschriebenes Deutsch und gesprochenes Deutsch. In *Von deutscher Sprache: Aufsätze, Vorträge und Plaudereien* (pp. 11–34). (Opera originale pubblicata 1899).
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Blake, C., & Klimmt, C. (2010). Geschlechtergerechte Formulierungen in Nachrichtentexten. *Publizistik*, 55(3), 289–304. <https://doi.org/10.1007/s11616-010-0093-2>.
- Bordin, B. (2022). *Gendergerechtigkeit in der Translationswissenschaft: Umgang mit genderinklusiver Sprache beim Dolmetschen im Deutschen und im Italienischen* (Tesi di laurea magistrale, Universität Wien). <https://doi.org/10.25365/thesis.72904>.
- Braun, F., Oelkers, S., Rogalski, K., Bosak, J., & Sczesny, S. (2007). «Aus Gründen della Verständlichkeit ...»: Der Einfluss generisch maskuliner und alternativer

Personenbezeichnungen auf die kognitive Verarbeitung von Texten. *Psychologische Rundschau*, 58(3), 183–189.

Bubenhofer, N. (2009). *Sprachgebrauchsmuster: Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. De Gruyter.

Bubenhofer, N. (2013). Quantitativ informierte qualitative Diskursanalyse: Korpuslinguistische Zugänge zu Einzeltexten und Serien. In K. Roth & C. Spiegel (A cura di), *Angewandte Diskurslinguistik: Felder, Probleme, Perspektiven* (pp. 109–134). Akademie Verlag.

Bürgerliches Gesetzbuch (BGB), § 611b (*non più in vigore*) (Germania).

Burnett, H., & Pozniak, C. (2021). Political dimensions of gender-inclusive writing in Parisian universities. *Journal of Sociolinguistics*, 25(5), 808–831. <https://doi.org/10.1111/josl.12489>

Calamai, S. (2023). La trascrizione automatica. In *Quaderni della Rassegna degli Archivi di Stato: Vademecum per il trattamento delle fonti orali* (pp. 219–226). Ministero della Cultura.

Cancelleria federale austriaca. (2024). *Geschlechtergerechte Sprache*. Bundeskanzleramt Österreich.

Cancelleria federale svizzera. (1996). *Leitfaden zur sprachlichen Gleichbehandlung*. Schweizerische Bundeskanzlei.

Cancelleria federale svizzera. (2023). *Geschlechtergerechte Sprache: Leitfaden zum geschlechtergerechten Formulieren in deutschsprachigen Texten des Bundes* (3^a ed. riveduta). Schweizerische Bundeskanzlei.

Chalyvidou, D., & Weber, A. (2025). Processing the gender star form in German: A comparison of written and spoken modalities. *Journal of Language and Social Psychology*. <https://doi.org/10.1177/0261927x251393827>

Corte costituzionale federale tedesca (Bundesverfassungsgericht). (2017, 10 ottobre). *Decisione in I BvR 2019/16 (sentenza)*. https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2017/10/rs201710_10_1bvr201916.html

De l'Horizon, K. (2022). *Blutbuch*. DuMont.

Diewald, G., & Steinhauer, A. (2022). *Handbuch geschlechtergerechte Sprache: Wie Sie angemessen und verständlich gendern* (2^a ed.). Dudenverlag.

Draxler, C. (2022). *Automatic transcription of spoken language using publicly available web services*. In *Fare linguistica applicata con le digital humanities*.

Ehlich, K., & Rehbein, J. (1976). Halbinterpretative Arbeitstranskriptionen (HIAT). *Linguistische Berichte*, 45, 21–41.

Ellsäßer, S. (2023). In the periphery of an indefinite pronoun: Forms and functions of conceptual agreement with jemand. *Zeitschrift für Sprachwissenschaft*, 42(2), 209–236.

Elmiger, D. (2009). Sprachliche Gleichbehandlung von Frau und Mann: Eine korpusgestützte Untersuchung über den Sprachwandel in der Schweiz. *Linguistik Online*, 39(3).

Elmiger, D. (2011). Von Dozierenden und Emeritierenden: Substantivierte Partizip-I-Formen im heutigen Deutsch. *Travaux Neuchâtelois de Linguistique*, 63, 163–179.

Elmiger, D., Tunger, V., & Schaeffer-Lacroix, E. (2017a). *Geschlechtergerechte Behördentexte: Linguistische Untersuchungen und Stimmen zur Umsetzung in der püsprachigen Schweiz* (FNS Project No. 143585). Université de Genève.

Elmiger, D., Tunger, V., & Schaeffer-Lacroix, E. (2017b). Geschlechtergerechte Sprache in Schweizer Behördentexten: Möglichkeiten und Grenzen einer püsprachigen Umsetzung. *OBST – Osnabrücker Beiträge zur Sprachtheorie*, 90, 61–90.

Erfurt, J. (1988). Feministische Sprachpolitik und soziolinguistische Aspekte des Sprachwandels. *Zeitschrift für Germanistik*, 9(6), 706–716.

Fiehler, R. (2009). Gesprochene Sprache. In *Duden. Die Grammatik* (8^a ed., pp. 1165–1244). Dudenverlag.

Formanowicz, M., Bedynska, S., Cislak, A., Braun, F., & Sczesny, S. (2013). Side effects of gender-fair language: How feminine job titles influence the evaluation of female applicants. *European Journal of Social Psychology*, 43(1), 62–71.

- Friedrich, M. C., & Heise, E. (2019). Does the use of gender-fair language influence the comprehensibility of texts? *Swiss Journal of Psychology*, 78(1–2), 51–60.
- Friedrich, M. C., Muselick, J., & Heise, E. (2022). Does the use of gender-fair language impair the comprehensibility of video lectures? *Psychology Learning & Teaching*, 21(3), 296–309.
- Glim, S., Körner, A., & Rummer, R. (2025). An ERP-based comparison of gender representations elicited by generic masculine role nouns and the German gender star form. *Journal of Neurolinguistics*, 73, 101231.
- Guentherodt, I., Hellinger, M., Pusch, L. F., & Trömel-Plötz, S. (1980). Richtlinien zur Vermeidung sexistischen Sprachgebrauchs. *Linguistische Berichte*, 69, 15–21.
- Gustafsson Sendén, M., Bäck, E. A., & Lindqvist, A. (2015). Introducing a gender-neutral pronoun in a natural gender language: The influence of time on attitudes and behavior. *Frontiers in Psychology*, 6, 893.
- Gustafsson Sendén, M., Renström, E., & Lindqvist, A. (2021). Pronouns beyond the binary: The change of attitudes and use over time. *Gender & Society*, 35, 588–615.
- Gygax, P., Gabriel, U., Sarasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted. *Language and Cognitive Processes*, 23(3), 464–485.
- Haß-Zumkehr, U. (2003). Hat die Frauenbewegung Wortschatzgeschichte geschrieben? In M. Wengeler (A cura di), *Deutsche Sprachgeschichte nach 1945* (pp. 161–179). Olms.
- Heinrich-Heine-Universität Düsseldorf. (2023). *Geschlechtergerechte Sprache: Eine Handreichung*.
- Hellinger, M. (A cura di). (1985). *Sprachwandel und feministische Sprachpolitik: Internationale Perspektiven*. VS Verlag.
- Hellinger, M., & Bierbach, C. (1993). *Eine Sprache für beide Geschlechter*. Deutsche UNESCO-Kommission.
- Hellinger, M., & Bußmann, H. (2001). *Gender across languages* (Vol. 3). John Benjamins.

Hennig, M. (2000). Grammatik der gesprochenen Sprache. *Informationen Deutsch als Fremdsprache*, 27(2–3), 218–220.

Hodel, L., Formanowicz, M., Pietraszkiewicz, A., & Sczesny, S. (2025). The role of social class in the use of gender-inclusive language: An analysis of Polish and German job titles. *Social Psychological Bulletin*.

Hodel, L., Formanowicz, M., Sczesny, S., Valdrova, J., & von Stockhausen, L. (2017). Gender-inclusive language in job advertisements. *Journal of Cross-Cultural Psychology*, 48(3), 384–401.

Hoffmann, L. (1998). *Grammatik della gesprochenen Sprache*. Stauffenburg.

Imo, W., & Lanwer, J. P. (2016). Sprache ist komplex. Nur: Für wen? In M. Hennig (A cura di), *Linguistische Komplexität – ein Phantom?* (pp. 149–175). Stauffenburg.

Ivanov, C., & Lieboldt, S. (2023). Von „Studierenden“, „Mitarbeiter*innen“ und „Professoren“: Geschlechtergerechte Personenbezeichnungen zwischen Varianz und Normierung. *Linguistik Online*.

Jöckel, S., Dogruel, L., & Bachofer, R. (2021). Effects of gender-fair addressing. *Publizistik*, 66, 441–462.

Judith, V. J. Z. (2025). Promoting global gender-fair language in college English classrooms. *Edelweiss Applied Science and Technology*, 9(5), 1893–1909.

Kalverkämper, H. (1979). Die Frauen und die Sprache. *Linguistische Berichte*, 62, 55–71.

Key, M. R. (1975). *Male/female language: With a comprehensive bibliography*. Scarecrow Press.

Klann-Delius, G. (2005). *Sprache und Geschlecht: Eine Einführung*. Metzler.

Klein, J. (1988). Die Benachteiligung der Frau im generischen Maskulinum – eine feministische Schimäre oder psycholinguistische Realität? In N. Oellers (a cura di), *Vorträge des Germanistentages Berlin 1987* (Vol. 1, *Das Selbstverständnis der Germanistik*).Tubingen: Niemeyer, 310–319.

- Kolek, V. (2025). Personal references on the front pages of the German and Austrian press. *Slovenščina 2.0*, 13(1), 1–23.
- Kolvenbach, M. (2019). Was kann das Wörtchen «man» dafür ...? *Sprachreport*, 3(4), 16.
- Kopf, K. (2022). Ist Sharon Manager? Anglizismen und das generische Maskulinum. In G. Diewald & D. Nübling (A cura di), *Genus – Sexus – Gender* (pp. 5–103). De Gruyter.
- Kopf, K. (2023). Normalfall Movierung: Geschichte und Gegenwart des generischen Maskulinums in Prädikativkonstruktionen. In D. Hetjens et al. (A cura di), *Historische (Morpho-)Syntax des Deutschen* (pp. 96–226). De Gruyter.
- Körner, A., Glim, S., & Rummer, R. (2024). Examining the glottal stop as a mark of gender-inclusive language in German. *Applied Psycholinguistics*, 45(1), 156–179.
- Kotthoff, H., & Nübling, D. (2024). *Genderlinguistik: Eine Einführung in Sprache, Gespräch und Geschlecht*. Narr Francke Attempto.
- Lakoff, R. (1973). Language and woman's place. *Language in Society*, 1(2), 45–80.
- Lange, E., & von Stockhausen, L. (2025). What affects people's use of and attitudes toward gender-inclusive language? *Frontiers in Psychology*, 16.
- Leech, G. (1997). Introducing corpus annotation. In R. Garside et al. (A cura di), *Corpus annotation* (pp. 1–18). Longman.
- Legge Fondamentale per la Repubblica Federale di Germania (Grundgesetz für die Bundesrepublik Deutschland). (1949). Art. 3, comma 2. https://www.gesetze-im-internet.de/gg/art_3.html
- Leibniz-Institut für Deutsche Sprache. (2024). Genus und Sexus. In *Grammatisches Informationssystem (GRAMMIS)*. <https://grammis.ids-mannheim.de/systematische-grammatik/2276>
- Linell, P. (2005). *The written language bias in linguistics: Its nature, origins and transformations*. Routledge.

- Link, S. (2024). The use of gender-fair language in Austria, Germany, and Switzerland: A contrastive, corpus-based study. *Lingua*.
- Miller, C., & Swift, K. (1972, 16 aprile). One small step for Genkind. *The New York Times*.
- Miller, C., & Swift, K. (2001). *The handbook of nonsexist writing* (2^a ed.). iUniverse.
- Müller-Spitzer, C., Ochs, S., Koplenig, A., Rüdiger, J. O., & Wolfer, S. (2024). Less than one percent of words would be affected by gender-inclusive language in German press texts. *Humanities and Social Sciences Communications*, 11, 1343.
- Nardone, C. (2018). *Lingua, genere e lavoro in Italia e in Germania* [Tesi di dottorato, Università di Bologna].
- Nübling, D., & Rosar, A. (2025). *Ist sie Anfänger oder Anfängerin? Eine korpuslinguistische Untersuchung*. De Gruyter. <https://doi.org/10.1515/9783111485102-009>
- Pabst, L., & Kollmayer, M. (2023). How to make a difference: The impact of gender-fair language on text comprehensibility. *Frontiers in Psychology*, 14.
- Parlamento tedesco. (1990). *Maskuline und feminine Personenbezeichnungen in der Rechtssprache* (Drucksache 12/1041).
- Preseau, L. (2025). A comparative analysis of gender-just language in first-year German textbooks. *Die Unterrichtspraxis/Teaching German*.
- Pusch, L. F. (1979). Der Mensch ist ein Gewohnheitstier. *Linguistische Berichte*, 63, 84–102.
- Pusch, L. F. (1984). *Das Deutsche als Männersprache*. Suhrkamp.
- Pusch, L. F. (2022). *Gegen das Schweigen: Meine etwas andere Kindheit und Jugend*. Wallstein.
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust speech recognition via large-scale weak supervision*. *arXiv*. <https://arxiv.org/abs/2212.04356>.
- Sacks, H., Schegloff, E.A. & Jefferson, G.: 1974, 'A simplest systematics for the organization of turn—taking for conversation', *Language* 50, 696–735.

- Samel, I. (2000). *Einführung in die feministische Sprachwissenschaft* (2^a ed.). Erich Schmidt.
- Sarrasin, O., Gabriel, U., & Gygax, P. (2012). Sexism and attitudes toward gender-neutral language. *Swiss Journal of Psychology*, 71, 113–124.
- Schlobinski, P. (A cura di). (1997). *Syntax des gesprochenen Deutsch*. VS Verlag.
- Schoenthal, G. (1989). Personenbezeichnungen im Deutschen als Gegenstand feministischer Sprachkritik. *Zeitschrift für germanistische Linguistik*, 17(3), 296–314.
- Schwitalla, J. (1997). *Gesprochenes Deutsch: Eine Einführung*. Tübingen, Germany: Max Niemeyer Verlag.
- Schwitalla, J. (2003). *Gesprochenes Deutsch: Eine Einführung*. Erich Schmidt Verlag.
- Sczesny, S., Formanowicz, M., & Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? *Journal of Social Issues*, 72(3), 488–512.
- Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., Birkner, K., ... Uhmann, S. (2009). *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)*. Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion, 10, 353–402.
- Sorbelli, S. (2024). *Linguaggio di genere e interpretazione: Un connubio possibile?* [Tesi di laurea magistrale non pubblicata]. Università di Bologna.
- Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the sexes in language. In K. Fiedler (A cura di), *Social communication* (pp. 163–187). Psychology Press.
- Stahlberg, D., Sczesny, S., & Braun, F. (2001). Name your favorite musician. *Journal of Language and Social Psychology*, 20(4), 464–469.
- Stecker, C., Müller, J., Blätte, A., & Leonhardt, C. (2021). *The evolution of gender-inclusive language: Evidence from the German Bundestag, 1949–2021* (Working paper).
- Storjohann, P. (2004). Frau auf dem linguistischen Prüfstand. *German Life and Letters*, 57(3), 309–330.

Straniero Sergio, F., & Falbo, C. (2012). Studying interpreting through corpora. In *Breaking ground in corpus-based interpreting studies* (pp. 9–52). Peter Lang.

Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10(1), 1–13.

Trömel-Plötz, S. (1978). Linguistik und Frauensprache. *Linguistische Berichte*, 57, 49–68.

Universität zu Köln, Gleichstellungsbeauftragte. (2021). *Überzeugendere Sprache: Leitfaden für eine geschlechtersensible Sprache* (7^a ed.).

Vervecken, D., & Hannover, B. (2015). Yes I can! Effects of gender fair job descriptions. *Social Psychology*, 46(2), 76–92.

Waldendorf, A. (2024). Words of change: The increase of gender-inclusive language in German media. *European Sociological Review*, 40(2), 357–374.

Wetschanow, K. (2017). Von nicht-sexistischem Sprachgebrauch zu fairen W_ortungen. *Osnabrücker Beiträge zur Sprachtheorie (OBST)*, 90, 33–60.

Wodak, R., Moosmüller, S., Feistritz, G., & Doleschal, U. (in stampa). *Frau und Mann im öffentlichen Bereich. Linguistische Richtlinien zur geschlechtsneutralen Sprachverwendung*. Wien.

Zifonun, G. (2000). Man lebt nur einmal. Morphosyntax und Semantik des Pronomens man. *Deutsche Sprache*, 28, 232–253.

Zifonun, G. (2018). Die demokratische Pflicht und das Sprachssystem. *Sprachreport*, 34, 44–56.

Sitografia

Archivio federale svizzero. (2025). *Pubblicazioni ufficiali e altri documenti digitalizzati*. <https://www.bar.admin.ch/bar/it/home/ricerca/ricerca/motori-di-ricerca-e-portali/pubblicazioni-ufficiali.html> (Consultato il 01.02.2026).

Baviera (Land). (s.d.). *Allgemeine Geschäftsordnung für die Behörden des Freistaates Bayern (AGO)*, § 22. <https://www.gesetze-bayern.de/Content/Document/BayAGO-22> (Consultato il 01.02.2026).

Dehoff-Zuch, J. (2016). *Leitfaden geschlechtergerechte Sprache: Eine Information della Gleichstellungsstelle*. Hochschule Emden/Leer. (Consultato il 02.02.2026).

Digitales Wörterbuch della deutschen Sprache. (s.d.-a). *DWDS-Korpus: Politische Reden (1982-2020)*. https://www.dwds.de/d/korpora/politische_reden (Consultato il 16.01.2026).

Digitales Wörterbuch der deutschen Sprache. (s.d.-b). *Wähler*. In *DWDS*. <https://www.dwds.de/wb/Wähler> (Consultato il 16.01.2026).

Duden. (s.d.). „Liebe Mitgliederinnen und Mitglieder“? – Personenbezeichnungen mit festem Genus. <https://www.duden.de/sprachwissen/> (Consultato il 29.01.2026).

FemBio. (s.d.). *FemBio – Frauen-Biographieforschung*. <https://www.fembio.org/> (Consultato il 21.02.2026).

Gendern. Portal Wissenschaftliches Schreiben. (s.d.). *Universität Leipzig Schreibportal*. <https://home.uni-leipzig.de/schreibportal/gendern/> (Consultato il 28.01.2026).

Gesellschaft für deutsche Sprache e. V. (GfdS). (2020). *Leitlinien zu den Möglichkeiten des Genderings*. <https://gfds.de/standpunkt-der-gfds-zu-einer-geschlechtergerechten-sprache/> (Consultato il 28.01.2026).

Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion. (2009). *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)* [PDF]. <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf>

HIAT-2 – EXMARaLDA. (s.d.). *HIAT-2: Halbinterpretative Arbeitstranskriptionen 2*. <https://exmaralda.org/en/hiat-2/> (Consultato il 21.02.2026)

Istituto dell'Enciclopedia Italiana. (s.d.). Corpus. In *Vocabolario Treccani online*. <https://www.treccani.it/vocabolario/genere/> (Consultato il 09.01.2026).

Olderdissen, C. (2019, 21 giugno). *Luise F. Pusch – Die feministische Linguistik hat ihr Leben bestimmt*. genderleicht.de. <https://www.genderleicht.de/> (Consultato il 28.01.2026).

OpenAI. (s.d.). *Whisper*. <https://openai.com/it-IT/index/whisper/> (Consultato il 10.01.2026).

Pusch, L. F. (2025). *Biografie*. luisepusch.de. <https://luisepusch.de/biografie/> (Recuperato il 10.01.2026).

pyannoteAI. (s.d.). *Welcome to pyannoteAI*. <https://docs.pyannote.ai/> (Consultato il 10.01.2026).

Sketch Engine. (s.d.). *User Guide / CQL basics / German RFTagger part-of-speech tagset / FAQ*. <https://www.sketchengine.eu/> (Consultato il 10.01.2026).

Wegner, J., & Amend, C. (2022, 22 novembre). *Luise Pusch, warum ist Deutsch eine Männersprache?* (Episodio 59) [Audio podcast episode]. In *Alles gesagt? Die Zeit*. <https://www.zeit.de/gesellschaft/2022-11/luisse-pusch-gendern-deutschland-interviewpodcast-alles-gesagt> (Consultato il 10.02.2026).

WhisperAI. (s.d.). *Frequently Asked Questions (FAQ)*. <https://whisperai.com/faq> (Consultato il 10.01.2026).

Appendice: Un nuovo approccio alla trascrizione automatica del parlato dialogico

La trascrizione e *diarizzazione* del parlato dialogico supportata da tecnologie di IA

di Gabriele Carioli

Introduzione – Il progetto

La fase di trascrizione ha richiesto un grande lavoro di esplorazione e sperimentazione delle diverse possibilità di automazione del processo, che altrimenti avrebbe richiesto delle tempistiche incompatibili con i limiti di tempo legati alla stesura del presente elaborato. Era infatti necessario trovare strumenti in grado non solo di trascrivere automaticamente un podcast di lunga durata (tre ore o più), ma anche di identificare correttamente i vari parlanti, separando i loro turni di parola e producendo delle trascrizioni grezze che fossero già abbastanza precise, in modo da minimizzare i tempi di revisione.

La durata dell'episodio (circa cinque ore e mezza), la presenza di diversi oratori e le frequenti sovrapposizioni tra gli stessi hanno rappresentato delle complessità aggiuntive, che abbiamo cercato di arginare implementando diverse soluzioni.

Riassumendo, gli obiettivi erano dunque i seguenti:

- Trascrivere podcast di lunga durata mantenendo la separazione degli speaker
- Minimizzare l'intervento manuale mantenendo alta qualità
- Produrre output multipli per i diversi casi d'uso
- Gestire le limitazioni tecniche dei modelli disponibili

Come prima cosa sono quindi stati individuati gli strumenti che ci avrebbero permesso di *diarizzare*, ovvero di identificare correttamente gli oratori, e di trascriverne i rispettivi turni di parola.

Oggi esistono molti modelli linguistici (LM) adatti a questo scopo, addestrati su decine di migliaia di ore di parlato. Tuttavia, i più performanti sono spesso servizi commerciali chiusi, che girano su datacenter remoti dotati di gruppi di schede grafiche potenti (GPU) e sono

accessibili tramite servizi a pagamento (API). Se da un lato questo permette a chiunque di usare una potenza di calcolo altrimenti inarrivabile per un dipartimento universitario o una piccola-media impresa, dall'altro comporta costi non trascurabili e nessuna garanzia sulla futura disponibilità del servizio o sulla stabilità delle sue interfacce di programmazione.

L'alternativa pratica è quindi utilizzare modelli open source. Per farlo in modo efficiente è però necessario un hardware dedicato, come workstation equipaggiate di GPU adeguate.

Il nostro Dipartimento di Interpretazione e Traduzione dispone di alcune workstation di buon livello, dotate di GPU Nvidia RTX 5000 con 32 GB di memoria dedicata, in grado di eseguire questi modelli.

Tra i LM open source disponibili, *Whisper* e *PyAnnote* si distinguono per essere i più maturi ed efficaci. Entrambi vantano un'ottima documentazione, dimensioni compatibili con il nostro hardware e, soprattutto, sono gratuiti e distribuibili con licenza Open Source.

***Whisper* e *PyAnnote*, punti di forza e criticità**

Whisper è un set di modelli di intelligenza artificiale per il riconoscimento vocale automatico sviluppato da OpenAI. È un sistema di trascrizione e traduzione *speech-to-text* di approccio *end-to-end*, progettato cioè per convertire direttamente il parlato in testo scritto, offrendo capacità robuste di trascrizione in numerose lingue e di traduzione diretta in inglese, il tutto all'interno di un'unica architettura modellistica coesa.

PyAnnote è un toolkit open-source, anch'esso composto di vari modelli, specializzato nell'elaborazione e nell'analisi del segnale audio parlato. La sua funzione principale e più riconosciuta è la *diarizzazione*, il processo computazionale che segmenta un flusso audio in base all'identità del parlante, rispondendo quindi alla domanda “chi parla quando?”. Fornisce un ecosistema di strumenti per la costruzione di pipeline applicative in questo specifico dominio.

La sfida principale è stata combinare queste due tecnologie di IA – *diarizzazione* e trascrizione *voice-to-text* – che presentano limitazioni complementari quando applicate ad audio estesi.

PyAnnote ha mostrato dei limiti nella gestione di audio molto lunghi, provocando la degradazione progressiva della qualità della *diarizzazione*. Questo è probabilmente anche dovuto al fatto che nell'arco di cinque ore sia il tono di voce che il registro tendono naturalmente a variare.

Whisper, il motore di trascrizione, lavora meglio con blocchi di audio di lunghezza moderata e coerenti; peggiora con segmenti troppo brevi o frammentati e tende ad avere “allucinazioni” con audio troppo lunghi.

A parte la scelta, più o meno obbligata, di questi modelli, le decisioni progettuali sono state determinate anche dalla complessità nel dover gestire questo tipo di nuove tecnologie, che abbiamo utilizzato per la prima volta direttamente, nonché dal fatto che una supervisione con interventi di correzione è inevitabile. Un ulteriore grado di complessità risiedeva nella necessità di consentire l’accesso diretto dell’hardware dedicato (workstation con GPU) alla laureanda, per utilizzare in modo estensivo l’applicazione.

Si è quindi evitata la creazione di un’applicazione monolitica, che avrebbe richiesto tempi di sviluppo molto più lunghi, e si è optato piuttosto per un’organizzazione del lavoro sotto forma di una pipeline modulare di operazioni “semplici” successive. Questo approccio affronta le limitazioni dei modelli attraverso un processo di segmentazione, filtraggio progressivo e ricostruzione temporale, permettendo all’operatore di intervenire manualmente solo dove strettamente necessario.

Scelte tecnologiche

PyAnnote Audio 4.0 con modello *speaker-diarization-community-1*

È stato scelto perché è uno dei pochi modelli di *diarizzazione* liberamente disponibili e sufficientemente accurati per uso non commerciale. Il recente modello *speaker-diarization-community-1*, rilasciato mentre lo sviluppo del progetto era in corso, si è rivelato essere più preciso del precedente *speaker-diarization-3.1* ed è stato pertanto subito adottato, comportando una riscrittura parziale del codice.

Il modello *diarizza* l’audio in quattro passaggi:

- *segmentation*: l’audio viene diviso in segmenti, rilevando l’attività vocale;
- *speaker_counting*: stima quanti sono i possibili speaker diversi;
- *embeddings*: vengono raccolte le caratteristiche vocali di ciascun segmento. Una voce viene rappresentata da 256 parametri numerici (dimensioni) che catturano timbro, altezza tonale, ritmo del parlato ed altre caratteristiche vocali uniche;
- *discrete_diarization*: fase di *clustering* in cui i segmenti vengono raggruppati per similarità (*clusterizzati*) in un numero di gruppi pari agli speaker trovati. I

segmenti in ciascun gruppo vengono attribuiti agli speaker con diversi gradi di sicurezza.

Gli *embeddings* vocali consolidati per ciascuno speaker, calcolati in base ai precedenti, vengono usati anche nella funzione sperimentale di mappatura automatica.

Modalità *Exclusive Mode* in *PyAnnote*

Sia *Whisper* che *Pyannote* hanno mostrato difficoltà ad elaborare i segmenti di audio che presentano sovrapposizioni vocali. Per far fronte a questa problematica si è optato per la modalità *Exclusive Mode* in *PyAnnote*, che assegna le parti sovrapposte solo allo speaker prevalente invece che a tutti gli speaker coinvolti, riducendo la frammentazione e migliorando di conseguenza la qualità dei segmenti per *Whisper*.

***Whisper* di OpenAI**

Selezionato per la sua buona accuratezza nella trascrizione multilingue e perché è disponibile come open source. Supporta particolarmente bene il tedesco, lingua target della nostra analisi. Il modello scelto è quello più avanzato: *large-v3*.

Whisper è un sistema di riconoscimento vocale *end-to-end* che processa l'audio attraverso un'architettura *encoder-decoder*:

- Encoder: converte l'input audio in una rappresentazione latente attraverso una rete neurale convoluzionale⁸⁰, estraendo features spettrali e temporali;
- Decoder: trasforma la rappresentazione latente⁸¹ in testo attraverso un transformer, generando token linguistici con meccanismi di *attention*.⁸²

Il modello supporta un centinaio di lingue con rilevamento automatico della lingua e può eseguire trascrizione pura (*speech-to-text*), traduzione in inglese (per lingue non inglesi) e identificazione di segmenti con punteggi di confidenza temporali.

⁸⁰ Le reti neurali convoluzionali si ispirano al modo in cui la corteccia visiva animale elabora le informazioni. L'encoder in questo caso utilizza questo meccanismo per analizzare il segnale audio e individuare le caratteristiche più importanti, scartando il rumore inutile e concentrandosi sulle caratteristiche rilevanti per il parlato.

⁸¹ Ovvero una rappresentazione compressa dell'audio, che conserva solo le informazioni acustiche fondamentali ed è priva di dettagli superflui.

⁸² Il meccanismo di *attention* imita l'attenzione cognitiva umana e permette al modello di "concentrarsi" dinamicamente sulle parti più rilevanti dei dati di input mentre sta generando ogni nuova parola, scegliendo la sequenza più probabile in base al modello della lingua.

La versione *large-v3* (1550M parametri) offre il miglior bilanciamento tra accuratezza e risorse computazionali, con performance particolarmente elevate sulle lingue principali grazie al training su dataset diversificati che includono contenuti accademici, tecnici e colloquiali.

Pipeline a blocchi per audio lunghi

I primi tentativi di *diarizzazione* con PyAnnote hanno subito evidenziato l'insorgere di problemi ed errori nell'elaborazione di file di durata superiore a tre ore, in particolare per il modello *speaker-diarization-3.1*. Si è dunque optato per una segmentazione del file in blocchi di 30 minuti, che ha prodotto risultati migliori rispetto ai tentativi precedenti, riducendo il numero di false attribuzioni e segmenti persi.

Il nuovo modello *speaker-diarization-community-1* sembrerebbe avere meno difficoltà nella gestione di simili file, ma per confermarlo servirebbe una verifica più approfondita. Alternativamente, rimane sempre possibile decidere di usare un solo blocco.

Mappatura speaker semi-automatica e automatica

Nella mappatura semi-automatica, il programma genera dei campioni di audio per i singoli speaker rilevati nei singoli blocchi. L'operatore deve ascoltare i campioni e creare associazioni consistenti attraverso i blocchi.

È prevista anche un'opzione per la mappatura automatica degli speaker basata sugli *embeddings*. Ciononostante, una verifica dell'avvenuta mappatura automatica è comunque necessaria, anche solo per dare nomi significativi agli speaker.

Filtraggio a due stadi (merge + durata)

Per ridurre la frammentazione dei segmenti di parlato è stato implementato un filtraggio composto da due stadi:

Merge consecutivi: unisce segmenti dello stesso speaker separati da pause brevi (inferiori a 3 secondi), ricostruendo turni di parola completi invece di frammenti.

Filtro durata: agisce successivamente e rimuove i segmenti residui troppo brevi (meno di 0,75 secondi di durata) che sono tipicamente artefatti, respiri, o frammenti non trascrivibili (o comunque poco significativi). Il confronto con la traccia audio originale in fase di revisione ha rivelato che non è stato perso nessun segmento significativo di parlato.

Blocchi audio ottimali per *Whisper*

Il programma passa a *Whisper* i segmenti audio di ciascuno speaker, con un *padding* di silenzio di 0.5 secondi prima e dopo il segmento interessato: in questo modo, l'audio non parte o termina all'improvviso e vengono migliorate le chance di riconoscimento per segmenti brevi.

Whisper può produrre più segmenti di testo per ogni segmento audio. Lo script genera una struttura dati che contiene sia i segmenti di testo separati, con i relativi *timestamp*, che la loro versione aggregata, con le temporizzazioni dei segmenti audio originali.

Revisione delle trascrizioni tramite uno strumento per il sottotitolaggio

Esportare le trascrizioni in formato SRT consente di utilizzare strumenti di sottotitolaggio noti, maturi e ben supportati come *SubtitleEdit*. Questi strumenti consentono di revisionare la trascrizione e di correggere gli eventuali errori potendo disporre dell'audio sincronizzato.

Approccio per fasi

Il sistema funziona usando una *directory di progetto* – una cartella dedicata che contiene tutti i dati, le configurazioni e gli output per l'elaborazione del file audio.

I dati sono persistenti e ogni fase genera un output che diventa l'input della fase successiva. I file intermedi vengono mantenuti, permettendo di:

- Riavviare la pipeline da qualsiasi punto;
- Analizzare i risultati intermedi;
- Modificare parametri e rieseguire solo fasi successive.

Le scelte effettuate nelle varie fasi vengono mantenute in un file di configurazione centralizzato (*config.yaml*) consentendo il *versioning* delle configurazioni, riproducibilità dei risultati, modifica fine manuale dei parametri.

Le *directory di progetto* possono essere facilmente trasferite per le fasi di lavoro manuali o per il backup, e l'isolamento consente di lavorare in parallelo su più versioni dello stesso podcast o su podcast diversi. Lo stato stesso della directory di progetto documenta l'avanzamento dell'elaborazione.

Le fasi nel dettaglio

Preparazione e diarizzazione

L'audio originale viene convertito in WAV 16kHz mono, poi diviso in blocchi di 30 minuti. Ogni blocco viene processato indipendentemente da *PyAnnote*.

Per ogni speaker rilevato, vengono estratti campioni audio di 30 secondi dai segmenti più lunghi. L'operatore ascolta questi campioni e compila la mappa speaker associando blocco per blocco i nomi temporanei degli speaker (SPEAKER_1, SPEAKER_2,...) a nomi permanenti (Intervistatore, Ospite,...).

I valori sono modificabili tramite le opzioni del comando. Alternativamente è possibile usare un solo blocco o richiedere la mappatura automatica degli speaker.

Unificazione

I dati di tutti i blocchi vengono uniti applicando la mappatura creata. Gli ID segmento originali vengono preservati per tracciabilità.

Pulizia

Segmenti consecutivi dello stesso speaker con pause inferiori ai 3 secondi vengono uniti. Segmenti di lunghezza inferiore ai 0,75 secondi rimasti isolati vengono rimossi. In questa fase si ha la possibilità di eliminare o di unire speaker erroneamente riconosciuti come diversi. I valori di default sono modificabili tramite le opzioni del comando.

Estrazione audio combinato (Opzionale)

Vengono create tracce audio separate per ogni speaker, utili per analisi o elaborazioni successive.

Trascrizione

I segmenti filtrati vengono passati a *Whisper* per la trascrizione.

Esportazione

Vengono generati i file di trascrizione, in formato SRT e come file di testo con e senza *timestamp*, sia suddivisi per *speaker* che combinati.

Revisione

Le trascrizioni vengono esportate in formato SRT per poter essere successivamente revisionate e corrette utilizzando strumenti di sottotitolaggio noti, maturi e ben supportati come *SubtitleEdit*, che permettono di allineare la trascrizione con l'audio originale.

È possibile scegliere se esportare le trascrizioni aggregate e/o quelle separate, con la segmentazione di *Whisper*.

Output finali

Al termine della procedura vengono prodotti i seguenti file:

- File JSON con trascrizioni complete e metadati;
- File SRT per editing con *SubtitleEdit*;
- File di testo con *timestamp*;
- File di testo puliti;
- Report HTML per analisi della qualità della *diarizzazione*.

Limitazioni note

Le limitazioni principali sono quelle intrinseche dei LM utilizzati:

- Qualità dipendente dalla qualità audio originale;
- Performance non ottimale con accenti fortemente regionali;
- Difficoltà con speaker molto simili vocalmente;
- Gestione non perfetta di musica e effetti sonori.

Queste limitazioni influiscono sulla qualità della *diarizzazione* e della trascrizione, rendendo necessari un intervento manuale nella mappatura degli speaker e una successiva fase di revisione. La produzione dei sample vocali e la funzione di mappatura automatica possono però mitigare, se non proprio azzerare, l'impatto di questo aspetto. La qualità delle trascrizioni grezze prodotte, combinata con l'utilizzo di *SubtitleEdit* per la correzione delle stesse, ha semplificato notevolmente il lavoro di revisione. Il sistema adottato rappresenta buon compromesso tra automazione e qualità e mette in luce le nuove possibilità che le tecnologie di IA oggi offrono in termini di trascrizione del parlato.

Ringraziamenti

Si ringraziano la Prof.ssa Gabriele Dorothe Mack, relatrice, e la Prof.ssa Doris Anita Höhmann, correlatrice, per la disponibilità, la competenza e la costante attenzione dimostrate durante la stesura della tesi. Un sentito ringraziamento a Gabriele Carioli, per il prezioso supporto tecnico fornito, che ha reso possibile lo sviluppo del lavoro.