

# Linguistica dei corpora

Una nuova prospettiva teorico-metodologica  
per lo studio di morfologia e sintassi in ambito DaF



Carolina Flinz



Milano University Press



Carolina Flinz

# LINGUISTICA DEI CORPORA

Una nuova prospettiva teorico-metodologica  
per lo studio di morfologia e sintassi in ambito *DaF*

*Linguistica dei corpora: una nuova prospettiva teorico-metodologica per lo studio di morfologia e sintassi in ambito DaF* / Carolina Flinz. Milano: Milano University Press, 2024.

ISBN 979-12-55101-11-6 (print)

ISBN 979-12-55101-13-0 (PDF)

ISBN 979-12-55101-15-4 (EPUB)

DOI 10.54103/milanoup.165

Questo volume e, in genere, quando non diversamente indicato, le pubblicazioni di Milano University Press sono sottoposti a un processo di revisione esterno sotto la responsabilità del Comitato editoriale e del Comitato Scientifico della casa editrice. Le opere pubblicate vengono valutate e approvate dal Comitato editoriale e devono essere conformi alla politica di revisione tra pari, al codice etico e alle misure antiplagio espressi nelle Linee Guida per pubblicare su MilanoUP.

Le edizioni digitali dell'opera sono rilasciate con licenza Creative Commons Attribution 4.0 - CC-BY-ND, il cui testo integrale è disponibile all'URL: <https://creativecommons.org/licenses/by-nd/4.0/>



 Le edizioni digitali online sono pubblicate in Open Access su:  
<https://libri.unimi.it/index.php/milanoup>

© The Author(s), 2024

© Milano University Press per la presente edizione

Pubblicato da:  
Milano University Press  
Via Festa del Perdono 7 – 20122 Milano  
Sito web: <https://milanoup.unimi.it>  
e-mail: [redazione.milanoup@unimi.it](mailto:redazione.milanoup@unimi.it)

L'edizione cartacea del volume può essere ordinata in tutte le librerie fisiche e online ed è distribuita da Ledizioni (<https://www.ledizioni.it/>)

# Sommario

Introduzione	9
--------------	---

## PARTE I

### INTRODUZIONE ALLA LINGUISTICA DEI CORPORA

Capitolo 1	
Caratteristiche di un corpus	17
1.1 Corpus: definizione	17
1.2 Elementi di un corpus	20
1.3 Approcci di analisi di un corpus	27
Capitolo 2	
Corpora e analisi linguistica	29
2.1 Token, type e lemmi	29
2.2 Il procedimento di ricerca in un corpus	33
2.3 Risultati della ricerca	35
2.4 Liste di frequenza	38
2.5 Liste di <i>keywords</i> e di <i>n-grams</i>	41
2.6 Profilo di parola	44
2.7 Altre possibilità di analisi	49
Capitolo 3	
Corpora in lingua tedesca	51
3.1 Classificare i corpora	51
3.2 Principali raccolte di corpora esistenti per la lingua tedesca	56
3.2.1 <i>Das Deutsche Referenzkorpus (DeReKo)</i>	57
3.2.2 Il DWDS – <i>Digitales Wörterbuch der Deutschen Sprache</i>	61
3.2.3 La <i>Datenbank für Gesprochenes Deutsch (DGD)</i>	66
3.2.4 Altri corpora	69
Capitolo 4	
Costruire un corpus <i>ad hoc</i>	73
4.1 Considerazioni preliminari	73
4.2 Costruire un corpus <i>ad hoc</i> con <i>Sketch Engine</i>	74
4.2.1 Opzione 1	75
4.2.2 Opzione 2	78

PARTE II  
DESCRIZIONE DI FENOMENI MORFOLOGICI E SINTATTICI BASATI  
SULL'USO DI CORPORA E APPLICAZIONI

Capitolo 1	
Morfologia	83
1.1 Introduzione	83
1.1.1 La parola e i suoi elementi costitutivi	83
1.1.2 Costruzione delle parole	87
1.2 Morfologia con i corpora	104
1.2.1 Caso studio I: composti con <i>Klima</i> in un corpus di CMC costruito <i>ad hoc</i>	104
1.2.2 Caso studio II: il confisso <i>-gate</i> nel corpus <i>Wikipedia-de</i>	109
1.3 Attività ed esercizi	113
1.3.1 La parola e i suoi elementi costitutivi	113
1.3.2 Costruzione delle parole: la flessione	114
1.3.3 Costruzione delle parole: la formazione delle parole	115
Capitolo 2	
Sintassi	119
2.1 Introduzione	120
2.1.1 Classi di parola	120
2.1.2 Sintagmi	131
2.1.3 Funzioni sintattiche nella frase semplice	139
2.1.4 Ordine delle parole a livello di frase	147
2.1.5 Frasi principali e secondarie	149
2.2 Sintassi con i corpora	155
2.2.1 Caso studio I: l'elemento polifunzionale <i>ganz</i>	155
2.2.2 Caso studio II: avverbi pronominali in <i>wo-</i> in <i>German Web 2020</i> . Il caso di <i>worin</i>	159
2.3 Attività ed esercizi	164
2.3.1 Classi di parola	164
2.3.2 Sintagmi	168
2.3.3 Funzioni sintattiche nella frase semplice	169
2.3.4 Ordine delle parole a livello di frase	170
2.3.5 Frasi principali e secondarie	171
Nota conclusiva	173
Bibliografia	175

Alla mia famiglia,  
ad Aurora e Federico



# Introduzione

“Die heutige italienische germanistische Linguistik passt zunehmend ihre überlieferte Tätigkeit im DaF- und Übersetzungsbereich sowie ihre Beschäftigung mit Grammatik und Texten den neuesten Erkenntnissen der sprachwissenschaftlichen Theorien an, fächert dabei ihre Forschungstätigkeit zunehmend auf und bezieht auch nationalspezifische Fragestellungen mit ein. Neue Forschungsprojekte sind u. a. der Korpuslinguistik und der kognitiven wie interkulturellen Linguistik gewidmet, wobei oft eine deutsch-italienisch-kontrastive und/oder DaF-didaktisch zentrierte Perspektive eingebracht wird.” (Foschi Albert/Hepp 2010: 1696)

Il passo sopra riportato contribuisce a descrivere, a vantaggio della comunità scientifica internazionale cui si rivolge l'intero volume *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch* (Krumm/Fandrych/Hufeisen/Riemer 2010), la situazione della linguistica tedesca in Italia, disciplina all'epoca ancora molto giovane<sup>1</sup>. Alla cura per ambiti di studio tradizionali come la didattica del tedesco come lingua straniera (*Deutsch als Fremdsprache – DaF*), la traduzione, la grammatica e la dimensione testuale, si rilevano specificità nazionali anche nel crescente interesse per nuovi filoni di ricerca come la linguistica dei corpora, la linguistica cognitiva, la linguistica interculturale e la linguistica contrastiva.

A quasi quindici anni di distanza, ritengo che questa descrizione sia ancora di grande attualità. In particolare, la sinergia tra linguistica tedesca, ambito *DaF* e linguistica dei corpora può essere tuttora considerata una novità nel nostro settore scientifico disciplinare. Nella teoria e prassi didattica universitaria, le connessioni teorico-metodologiche tra le tre discipline sopra menzionate si realizzano infatti soprattutto a due a due: (1) tra linguistica dei corpora e linguistica tedesca, (2) tra linguistica tedesca e *DaF*, (3) tra linguistica dei corpora e *DaF*.

Il primo binomio (1) è attestato per ogni ambito della linguistica tedesca. Teoria e descrizione grammaticale (si veda per esempio il progetto *Korpusgrammatik* del *Leibniz-Institut für Deutsche Sprache*)<sup>2</sup>, come pure gli ambiti applicativi della disciplina, non possono fare a meno, della linguistica dei corpora (cfr. Beißwenger/Lemmitzer/Müller-Spitzer 2022). Per rispondere alle diverse questioni poste dalla ricerca, la maggior parte degli studi utilizza oggi dati autentici presenti nei vari tipi di corpora e gli strumenti per l'analisi messi a disposizione dai diversi sistemi di gestione dei corpora, focalizzando man mano singoli aspetti o cercando di cogliere, in base all'approccio scelto, una visione d'insieme del fenomeno studiato.

---

1 La denominazione della disciplina accademica *Lingua e Traduzione tedesca*, presente ancora tutt'oggi nella declaratoria del settore L-LIN/14, risale al 2001.

2 Cfr. <https://www.ids-mannheim.de/gra/projekte/korpusgrammatik/> [31.08.23].

Il secondo binomio (2) è molto diffuso nella prassi didattica universitaria, essendo pressoché indubbio che la riflessione (meta)linguistica serva a velocizzare e promuovere l'acquisizione di competenze linguistiche negli apprendenti di lingua straniera. Molte grammatiche didattiche del tedesco e manuali di introduzione alla linguistica tedesca mettono in evidenza questo aspetto, sottolineando come gli strumenti di analisi e di ricerca linguistica possano favorire l'apprendimento delle lingue straniere, la descrizione grammaticale, accelerare e sistematizzare la padronanza del tedesco come lingua straniera (cfr. Gaeta 2017: 14) e come una sistematica conoscenza metalinguistica possa facilitare l'apprendimento della lingua stessa (cfr. Di Meola 2014: 14).

Il terzo binomio (3) è invece ancora poco diffuso, nonostante le due discipline, linguistica dei corpora e *DaF*, siano accomunate dall'approccio empirico (cfr. Fandrych/Tschirner 2007). L'auspicio è che si giunga presto a una maggiore popolarità, poiché l'uso dei corpora, come dimostrato anche da studi empirici (cfr. Flinz 2021d), porta a un incremento dello *Sprachbewusstsein* negli apprendenti, favorendo l'apprendimento della lingua.

In un dialogo a tre fra linguistica dei corpora, linguistica tedesca e ambito *DaF*, sarebbe proprio la linguistica dei corpora a fare da *fil rouge*: è infatti all'interno di questa disciplina che possono essere condotte analisi in dati di vario tipo per studiare le caratteristiche della lingua e i fenomeni chiave della linguistica tedesca; sempre per mezzo della linguistica dei corpora l'oggetto di studio può essere compreso e appreso mediante una riflessione (meta)linguistica che osservi il funzionamento e lo sviluppo dei fenomeni grammaticali su dati autentici. Promuovendo consapevolezza di tecniche di rilevamento dei dati linguistici e strumenti di uso dei corpora aumenta infatti le possibilità di riflettere sulla lingua – non solo in base alla descrizione del suo funzionamento (apporto della linguistica), ma anche all'osservazione diretta di modalità, frequenze, varietà e mutamenti (tendenze e fenomeni storici di grammaticalizzazione). Il dialogo tra le discipline avviene proprio nel momento in cui si applica a contenuti linguistici di didattica *DaF* un orizzonte di studi di linguistica dei corpora.

L'esigenza, per la linguistica tedesca in Italia, di assimilare principi teorico-metodologici della linguistica dei corpora emerge anche nella prassi didattica. Come ho potuto constatare personalmente nel corso della mia esperienza come docente di linguistica tedesca presso l'Università degli Studi di Milano molte studentesse e molti studenti formati in ambito linguistico e letterario, pur utilizzando come testi di riferimento per lo studio universitario, testi specialistici fondati sull'uso di corpora, non possiedono le conoscenze e competenze di base utili a comprenderne il funzionamento e necessarie a consentirne un uso adeguato. Esperimenti didattici volti a colmare la lacunosa *corpus literacy*<sup>3</sup> di

---

3 Per *corpus literacy* si intende l'abilità di saper usare autonomamente i corpora (si veda Mukherjee 2002: 179-80).

studentesse e studenti (cfr. Mukherjee 2002) hanno messo altresì in evidenza come le competenze acquisite tramite l'uso della linguistica dei corpora possano rivelarsi efficaci anche per rafforzare la descrizione, comprensione e ricezione dei fenomeni linguistici e favorirne l'apprendimento.

Obiettivo del volume è articolare una proposta teorico-metodologica in cui far confluire le tre discipline, esemplificata su fenomeni scelti di morfologia e sintassi del tedesco, per la didattica *DaF* di ambito universitario.

I fenomeni individuati riguardano due livelli del sistema linguistico centrali per la descrizione grammaticale (cfr. Konopka 2020: 13), particolarmente adatti a essere affrontati e analizzati mediante la linguistica dei corpora. L'analisi di morfologia e sintassi qui presentata potrà servire da modello per altri livelli linguistici.

Il volume si compone di due parti. La PARTE I pone i fondamenti della linguistica dei corpora; il focus è sulla linguistica dei corpora di area tedescofona e italoфона. Alla presentazione e discussione di alcuni concetti base (capitolo 1) necessari per muovere i primi passi in questo ambito, seguono, nel capitolo 2, informazioni utili per l'analisi linguistica di un corpus. Metodi e approcci, procedure di analisi sono messi a confronto, per riflettere sulle loro specificità, sui loro vantaggi, ma anche sulle loro possibili criticità. Nel capitolo 3 dopo la presentazione e discussione di una proposta di classificazione dei corpora in base a caratteristiche specifiche, utile per orientarsi nel folto panorama delle risorse disponibili in rete, si presentano i principali corpora della lingua tedesca. Caratteristiche e funzionalità principali sono messe in evidenza. Nel capitolo 4 si forniscono le informazioni necessarie per la costruzione di un corpus *ad hoc* utilizzando il tool *Sketch Engine*. Si presentano diverse possibilità unitamente a riflessioni su eventuali applicazioni.

La PARTE II è costituita da una introduzione alla morfologia (capitolo 1) e alla sintassi del tedesco (capitolo 2). In ogni capitolo si parte da una sintesi del livello di studio selezionato, facendo perno su testi descrittivi grammaticali canonici, su grammatiche didattiche, su letture di approfondimento in testi specialistici e su testi di riferimento diffusi nell'insegnamento della linguistica tedesca nell'accademia italiana. Scopo della descrizione è procedere in modo integrativo, senza privilegiare un unico modello teorico rispetto ad altri, fornendo in tal modo un orientamento di base relativo ai principali meccanismi di funzionamento del tedesco in modo tale da garantire la padronanza dei concetti fondamentali necessari per procedere attivamente all'analisi linguistica di corpora. Segue poi l'esemplificazione del metodo di analisi a fenomeni paradigmatici (casi studio) all'interno dell'ambito selezionato:

- per la morfologia il focus è sulla formazione di parola. Obiettivo di analisi sono le parole composte con *Klima* in un corpus CMC costruito *ad hoc* (caso studio I) e il confisso *-gate* nel corpus *Wikipedia-de* (caso studio II).

- per la sintassi si è scelto di focalizzare l'attenzione sulle classi di parola, la cui identificazione è necessaria per la determinazione dei sintagmi. Nel primo caso studio (caso studio I) si prende in esame l'elemento polifunzionale *ganx* in due tipi di corpora diversi: il *web corpus German Web 2020* e il corpus di tedesco scritto DeReKo-2023-1. Nel secondo caso studio sono analizzate le forme (integrate, non integrate) degli avverbi pronominali in *wo* in DeReKo-2023-1.

Alla parte metodologica fa seguito la parte applicativa, con la presentazione di esercizi che prevedono l'uso di corpora e dei loro strumenti di analisi per l'identificazione e l'analisi di alcune caratteristiche specifiche della lingua tedesca, nonché della successiva riflessione a riguardo e la riflessione di alcune caratteristiche specifiche della lingua tedesca. Le attività applicative sono strettamente collegate alla parte teorica e metodologica, in modo da poter passare agevolmente dalla teoria alla pratica e viceversa. L'obiettivo generico è il rafforzamento delle conoscenze teoriche per mezzo dei corpora, con i quali la lingua può essere analizzata sia da un punto di vista induttivo che deduttivo. Al fine di sviluppare al meglio la *corpus literacy*, si è scelto di utilizzare in prevalenza il corpus *German Web 2020* presente in *Sketch Engine* e le funzionalità di questo sistema di analisi. Il sistema è infatti semplice e intuitivo e molte delle sue caratteristiche si ritrovano anche in altri *tool*. Per coloro che si avvicinano per la prima volta alla linguistica dei corpora sono fornite le stringhe di ricerca necessarie (anche, in alcuni casi, con espressioni regolari)<sup>4</sup> a scopo esemplificativo; ma la presenza di ulteriori input permette di svolgere anche ricerche in maniera più autonoma e ricorrendo anche all'uso di altri corpora.

Il volume si rivolge a coloro che per necessità di ricerca o di studio vogliono approfondire dal punto di vista teorico i concetti centrali della linguistica dei corpora applicati ai due ambiti della linguistica tedesca scelti a modello: la morfologia e la sintassi. Le conoscenze acquisite a livello teorico possono poi essere applicate per condurre analisi empiriche volte a comprendere i fenomeni linguistici oggetto della propria ricerca.

Il volume è pensato da un lato per docenti di linguistica tedesca e per docenti in formazione; al fine di insegnare in corsi in cui al centro vi sia la sinergia tra le tre discipline, è necessario, infatti, che essi abbiano sviluppato la *corpus literacy* adeguata. Dall'altro si rivolge anche a studentesse/studenti di germanistica e di tedesco come lingua straniera. Oltre che per corsi universitari tradizionali, il volume è pensato anche per l'auto-apprendimento. Studiare il tedesco a livello universitario implica l'acquisizione anche di metodologie di analisi e di studio, che permettano di elaborare riflessioni collegando le varie parti trattate nel volume e/o nei vari testi già studiati e nelle varie discipline, grazie anche all'uso

---

4 Per espressioni regolari si intende una sequenza di simboli che identifica in modo abbreviato una stringa di testo (cfr. Lemnitzer/Zinsmeister 2015: 92).

di una terminologia plurilingue, che permette di riconoscere termini consueti in altre aree disciplinari. Le riflessioni non si fermano però solo al livello intralinguistico, ma comprendono anche il livello interlinguistico; in alcuni casi sono presenti, pertanto, confronti con la lingua italiana.

Ringrazio l'Università degli Studi di Milano, il Dipartimento di Lingue, Letterature, Culture e Mediazioni nelle persone di Marina Brambilla e Marco Castellari, e la Milano University Press per avermi dato la possibilità di pubblicare questo volume. Desidero ringraziare anche tutte le persone che in diverso modo hanno contribuito alla realizzazione di questo progetto, in particolare i revisori anonimi e Sabrina Ballestracci, Marcella Costa, Marina Foschi Albert, Marianne Hepp, Simona Leonardi, Rita Luppi, Fabio Mollica, Manuela Moroni, Miriam Ravetto e Daniela Sorrentino per i preziosi consigli e commenti. Ringrazio infine tutte le mie studentesse e i miei studenti che durante i corsi, con le loro curiosità e domande, hanno contribuito alla stesura di questo volume.



PARTE I  
INTRODUZIONE ALLA LINGUISTICA  
DEI CORPORA



# Capitolo 1

## Caratteristiche di un corpus

In questo capitolo, dopo un breve accenno alla linguistica dei corpora<sup>1</sup> (*Korpuslinguistik*) e ai suoi esordi, si definisce cosa è un ‘corpus’ (paragrafo 1.1). Dalla definizione vengono estrapolati poi gli elementi chiave necessari per una corretta comprensione del termine: i dati primari (*Primärdaten*), i metadati (*Metadaten*) e le annotazioni (*Annotationen*) (paragrafo 1.2). Infine, il capitolo si conclude con una riflessione sui possibili approcci di analisi (paragrafo 1.3).

### 1.1 Corpus: definizione

La linguistica dei corpora (*Korpuslinguistik*), definita come lo studio della lingua basata su esempi di un uso linguistico reale<sup>2</sup> raccolti in un corpus (cfr. Stefanowitsch 2020: 22), ha una lunga tradizione, radicata nello strutturalismo e nell’empirismo (cfr. Rossini Favretti 2001: 368). I suoi esordi sono da collocare soprattutto nel mondo anglosassone, dove la lingua inglese è stata oggetto di numerose indagini<sup>3</sup>.

Contrapponendosi in modo netto all’introspezione del generativismo (cfr. Chomsky 1957), la linguistica dei corpora ha dato avvio a nuove prospettive di analisi (cfr. Firth 1957; McEnery/Wilson 1996/2001; Sinclair 1991), con applicazioni soprattutto in ambito lessicografico, grammaticale e didattico<sup>4</sup>. Il *Brown University Standard Corpus of Present-Day American English*, noto oggi soprattutto come *Brown Corpus*, pubblicato nel 1964 (cfr. Francis/Kučera 1964), è stato considerato per molto tempo uno standard da prendere come modello<sup>5</sup>. Composto

- 
- 1 Alcuni termini centrali verranno accompagnati dal corrispettivo tedesco, tra parentesi. Salvo diversa indicazione tutte le traduzioni sono dell’autrice.
  - 2 Cfr. “the study of language based on examples of ‘real life’ language use” (McEnery/Wilson 2001: 1).
  - 3 Inizialmente gli studi hanno interessato la lingua scritta (corpora di lingua scritta), successivamente quella parlata (corpora di lingua parlata). Solo in una fase successiva gli interessi si sono concentrati su altri aspetti, come l’acquisizione del linguaggio da parte di bambini in varie parti del mondo (progetto *CHILDES*), la variazione della lingua da un punto di vista diacronico (*Helsinki Corpus of English Texts: Diachronic Part*). Si veda Spina (2001: 61).
  - 4 Cfr. il progetto COBUILD (cfr. Sinclair 1987) che per la pubblicazione del Collins *COBUILD English Dictionary* si è avvalso di un grande corpus dell’inglese contemporaneo. In una fase successiva il progetto si è evoluto in una nuova iniziativa, *The Bank of English*, che negli anni è stata incrementata con l’aggiunta di nuovi testi.
  - 5 Corpora che hanno seguito i medesimi criteri per la loro formazione sono, tra gli altri, il *Lancaster-Oslo/Bergen Corpus* (LOB), il *British English*, il *Wellington Corpus of Written New Zealand English*, che sono stati definiti ‘corpora di prima generazione’, avendo dimensione ridotta rispetto ai corpora moderni; inoltre, si concentrano solo sulla variante scritta (cfr. Spina 2001:

da un milione di parole e formato da 500 campioni di 2.000 parole ciascuno, intendeva rappresentare, tramite vari tipi di testi (giornalistici e letterari), la varietà scritta dell'inglese d'America (cfr. Spina 2001: 58).

Sviluppi analoghi si sono verificati successivamente anche in altri paesi, tra cui l'Italia, dove grazie a una linguistica dei corpora *ante litteram*, ravvisabile nella linea empirica della tradizione lessicografica italiana (cfr. Sabatini in Barbera 2013: 15), sono state concepite opere quali quella di Roberto Busa su Tommaso d'Aquino (cfr. *Index Thomisticus* in Barbera 2013: 12) o il *Dizionario della Crusca*. L'interesse principale era qui rivolto al latino e all'italiano letterario delle origini o alla stesura di dizionari di frequenza dell'italiano contemporaneo per finalità didattiche (cfr. Spina 2001: 61)<sup>6</sup>.

In Germania gli sviluppi sono stati invece più recenti. Le prime introduzioni alla linguistica dei corpora risalgono al 2006 (cfr. Lemnitzer/Zinsmeister 2006; Scherer 2006), mentre l'uso dei metodi propri della disciplina per la linguistica tedesca è ancora successivo (cfr. Fandrych/Tschirner 2007). Nel 2022 è stato pubblicato il volume *Forschen in der Linguistik* (cfr. Beißwenger/Lemnitzer/Müller-Spitzer 2022), che raccoglie sia capitoli introduttivi relativi a singoli aspetti metodologici, a risorse digitali e a strumenti di analisi, sia casi studio. Per quanto riguarda il DaF (*Deutsch als Fremdsprache*), invece, a ora non è stata ancora pubblicata una introduzione analoga a quella per la didattica dell'inglese realizzata da Joybrato Mukherjee (2002). Per ovviare a questa lacuna, nel 2020 è stata fondata la rivista *Korpora Deutsch als Fremdsprache (KorDaF)*<sup>7</sup>, in cui non si pubblicano solo saggi di vario tipo, teorici, empirici, didattici, ma vengono presentati regolarmente anche singoli corpora e sistemi di analisi sotto forma di *portraits*.

Al di là di ogni datazione e percorso di sviluppo, la linguistica dei corpora riveste oggi un ruolo centrale per la linguistica in toto. La definizione di corpus (*das Korpus*) è ormai consolidata<sup>8</sup>:

Ein Korpus ist eine Sammlung schriftlicher oder gesprochener Äußerungen in einer oder mehreren Sprachen. Die Daten des Korpus sind digitalisiert, d.h. auf Rechnern gespeichert und maschinenlesbar (Lemnitzer/Zinsmeister 2015: 39).

---

60). Il primo corpus di riferimento della lingua inglese contemporanea, sia scritta sia parlata, è invece il *British National Corpus* (BNC), che nel 1994 constava di 100 milioni di parole, annotate grammaticalmente.

6 I primi dizionari di questo tipo si basavano su corpora, ma non li mettevano a disposizione. Cfr. <https://www.corpusthomicum.org/it/index.age> e <https://accademiadellacrusca.it/it/contenuti/dizionari/6225> [21.10.2022];

7 Cfr. <https://kordaf.tu-journals.ulb.tu-darmstadt.de> [21.10.2022]

8 Caratteristica non presente in questa definizione, ma in un certo senso presupposta, è l'autenticità dei dati (cfr. Spina 2001: 53), che devono essere esempi d'uso reale e non inventati artificialmente dal linguista (cfr. Freddi 2019: 11).

Fondamentale per questa definizione è in primis il formato, che deve essere digitale e pronto per essere elaborato da un computer; in caso contrario, non è possibile parlare di corpora, bensì solo di *Datensammlungen* (raccolte di dati)<sup>9</sup>. Il secondo aspetto sono i dati stessi, ossia le porzioni di lingua<sup>10</sup>, che possono essere di tipo scritto o parlato in una o più lingue.

Spina (2001: 64) aggiunge nella sua definizione di corpus le caratteristiche di rappresentatività e strutturazione: un corpus è quindi «una raccolta strutturata di testi in formato elettronico che si ritengono rappresentativi di una data lingua o di un suo sottoinsieme, mirata ad analisi di tipo linguistico».

Dal momento che non è possibile inserire in un corpus tutte le occorrenze testuali di una lingua in quanto sistema infinito, si deve procedere per campionamento (ossia per scelta e classificazione), seguendo criteri stabiliti in precedenza. Si dovrà cioè selezionare un numero sufficientemente elevato di dati e con una varietà tale da essere rappresentativa della lingua scritta e/o parlata oggi usata. Inoltre, occorre tenere conto della variazione diamesica (asse scritto-parlato), diafasica (genere e registro), diastratica (variazione sociale) e diatopica (variazione geografica), affinché il corpus rappresenti quanto più possibile le molteplici variazioni della lingua in questione. La rappresentazione, inoltre, dovrà essere anche bilanciata, per dare una fedele e strutturata raffigurazione della realtà (cfr. Freddi 2019: 11-13). Il criterio della rappresentatività viene soddisfatto in modo piuttosto significativo dai corpora di riferimento delle varie lingue, come *Das Deutsche Referenzkorpus* (DeReKo) per il tedesco<sup>11</sup>, il *British National Corpus* (BNC) per l'inglese britannico<sup>12</sup>, il *Corpus di Italiano Scritto* (CORIS/CODIS) per la lingua italiana<sup>13</sup>. Per le lingue europee è in fase di progettazione l'*Europäisches Referenzkorpus* (EuReCo) (cfr. Čermáková et al. 2021; Trawiński/Kupietz 2021).

Disponendo oggi di informazioni chiare ed esaurienti sui tipi di corpora esistenti e potendo realizzare facilmente corpora *ad hoc*, grazie allo sviluppo di software per la creazione, l'annotazione e l'analisi automatica, la rappresentatività è diventato un criterio non necessario per un corpus (si veda anche Lemnitzer 2022: 411).

9 Per la linguistica dei corpora lo scopo primario della costruzione di un corpus è l'analisi linguistica (cfr. Lemnitzer/Zinsmeister 2015: 40-41). Non si può fare a meno di evidenziare che anche corpora o archivi originariamente creati per altri scopi, oggi si prestano benissimo per analisi di tipo linguistico (cfr. tra gli altri il progetto *Gutenberg, Deutsches Textarchiv, Deutsche Wortwarte* ecc.). Anche il *World Wide Web*, previa riflessioni di tipo organizzativo e metodologico, può essere utilizzato per analisi linguistiche (cfr. Kilgarriff/Grefenstette 2003, Baroni/Bernardini 2006; Hundt et al. 2007).

10 Possono costituire un corpus sia porzioni di testo e di espressioni orali, sia testi ed espressioni orali nella loro interezza. Soprattutto in caso di problemi di copyright, è prassi comune usare frasi ordinate in maniera casuale (cfr. Lemnitzer/Zinsmeister 2015: 40).

11 Cfr. <https://www.ids-mannheim.de/digspra/kl/projekte/korpora/> [22.10.22].

12 Cfr. [www.natcorp.ox.ac.uk](http://www.natcorp.ox.ac.uk) [22.10.22].

13 Cfr. [https://corpora.ficlit.unibo.it/coris\\_ita.html](https://corpora.ficlit.unibo.it/coris_ita.html) [22.10.22].

## 1.2 Elementi di un corpus

Oltre ai suoi dati primari (*Primärdaten*), cioè i testi scritti e le espressioni orali che lo costituiscono, un corpus può contenere anche altri dati, come i metadati (*Metadaten*), vale a dire le informazioni sui dati stessi, e le annotazioni (*Annotationen*), ossia l'abbinamento dei dati con la loro descrizione:

Die Bestandteile des Korpus, die Texte oder Äußerungsfolgen, bestehen aus den Daten selber sowie möglicherweise aus Metadaten, die diese Daten beschreiben, und aus linguistischen Annotationen, die diesen Daten zugeordnet sind. (Lemnitzer/Zinsmeister 2015: 39)

I dati primari (*Primärdaten*) sono quindi i testi scritti e le espressioni orali, o loro porzioni, che compongono il corpus. Per la loro analisi, occorre tenere presente che alcune informazioni possono essere andate perdute nel passaggio dal testo originario al testo digitale (si pensi anche alla trasformazione di un testo tramite un software OCR)<sup>14</sup>, come per esempio la suddivisione in sillabe quando si va a capo, la distinzione tra testo centrale (*Kerntext*) e paratesto (*Paratext*), la suddivisione in paragrafi, l'evidenziazione grafica (grassetto, corsivo ecc.), le pause e le esitazioni, se si tratta di uno spezzone di lingua parlata ecc. Con questa codifica di primo grado molte delle caratteristiche individualizzanti vengono eliminate ai fini di una maggiore uniformità nella codifica dei dati<sup>15</sup> (cfr. Spina 2001: 74). Con la codifica di secondo grado ci sarà poi una normalizzazione dei testi, ossia la cosiddetta 'ripulitura' (correzione ortografica, eliminazione di righe vuote e spazi bianchi non necessari ecc.), fondamentale per la manipolazione e l'analisi dei dati.

I metadati (*Metadaten*), invece, sono le informazioni sui dati primari che descrivono alcuni degli aspetti principali, come per esempio il contenuto, l'autore, la casa editrice, l'arco temporale e/o il luogo in cui sono stati prodotti/raccolti, nonché altre informazioni bibliografiche. Per esempio, per il corpus *Emigrantendeutsch in Israel* nella *Datenbank für Gesprochenes Deutsch* sono presenti i seguenti metadati:

14 OCR sta per *Optical Character Recognition*. Con questi software è possibile rilevare i caratteri contenuti in un documento e convertirli in formati come ASCII (*Standard Code for Information Interchange*) o *Unicode*. Il risultato può essere poi modificato con un normale editor di testo. Cfr. [https://it.wikipedia.org/wiki/Riconoscimento\\_ottico\\_dei\\_caratteri](https://it.wikipedia.org/wiki/Riconoscimento_ottico_dei_caratteri) [21.11.22].

15 I file vengono salvati in formato solo testo. Uno dei primi sistemi di codifica è quello dei caratteri dello *Standard Code for Information Interchange* (ASCII) (cfr. Spina 2001: 75). Questo sistema, che si basa sull'alfabeto latino, ha mostrato i suoi limiti per altri tipi di alfabeti e per la lingua parlata, codificata con l'alfabeto della *International Phonetics Association* (IPA), che utilizza simboli e caratteri non ASCII. Nel corso degli anni sono stati sviluppati sistemi di codifica alternativi per il parlato (cfr. lo *Speech Assessment Methods Phonetic Alphabet*, ossia SAMPA), ma oggi il codice maggiormente in uso è il sistema *Unicode*.

- metadati che riguardano l'intervista (*Ereignis* = E): *DGD-Kennung, Sonstige Bezeichnung, in DGD seit Version, Aufnahmeort (Land), Aufnahmeort (Ortsname), Datum, Forscherin*;
- metadati che riguardano le persone intervistate (*Sprecher/Sprecherin* = S): *DGD-Kennung, Name, Früherer Name, Geschlecht, Geburtsdatum (Jahr), Geburtsort (Land), Geburtsort (Region), Geburtsort (Kreis)* ecc.

I metadati sono molto importanti per la documentazione e sono necessari per filtrare i dati primari in base a determinate informazioni, così come per la creazione di sottocorpora. Per la stesura di metadati esistono standard a cui riferirsi, come per esempio la *Dublin Core Metadata Initiative*, ampliata dalla *Open Language Archives Community* (OLAC) con informazioni specifiche per lingua<sup>16</sup>, il *Corpus Encoding Standard* (CES), che fa uso delle convenzioni usate per la *Text Encoding Initiative* (TEI), e la *Component Metadata Infrastructure* (CMDI), che permette anche la connessione tra metadati di varie risorse<sup>17</sup>.

L'annotazione (*Annotation*) è la procedura con la quale i dati primari, che si trovano ancora in uno stato grezzo (ossia non annotato), vengono etichettati in maniera sistematica con speciali etichette, per mettere in evidenza determinate informazioni. Si tratta, in sostanza, di una descrizione e interpretazione del testo, ovvero di «una sorta di metacodifica dei testi costituita da un insieme di simboli per etichettarne le diverse parti e di regole per riutilizzarli» (Spina 2001: 79).

L'annotazione può essere di vario tipo. Anche se annotazione e marcatura vengono spesso utilizzati come sinonimi (per esempio in Spina 2001), normalmente si preferisce differenziare tra *mark-up* (marcatura) e annotazione in senso stretto (cfr. Freddi 2019: 18-19).

Nel primo caso, nella marcatura, si codificano le informazioni oggettive relative ai dati, come per esempio le informazioni riguardanti la struttura del testo. Nell'esempio (1) si mostra una frase codificata nella *Standard Generalized Markup Language* (SGML)<sup>18</sup> dove le etichette *s* (*s* = inizio frase) e */s* (*/s* = fine frase) sono racchiuse dai simboli *<>*

- (1) *<s>*Die Annotationen werden auf mehreren Ebenen möglich sein, z. B. Verfasser, Übersetzer, Herausgeber, sodass Vergleiche vorgenommen werden können.*</s>* (German Web 2020)

Nel linguaggio SGML, per descrivere una porzione di testo si inserisce una marca, specificando sempre l'inizio e la fine (indicata con il simbolo */*). I *tag*,

16 Cfr. <http://www.language-archives.org> [03.12.23].

17 Per un approfondimento si vedano Lemnitzer/Zinsmeister (2015: 46-47).

18 La sua evoluzione è la *Extensible Markup Language* (EML), nata per permettere la distribuzione dei testi annotati tramite Internet.

ossia le etichette, possono essere attribuite a una parola singola, a una sequenza di parole o a una relazione (cfr. Lemnitzer/Zinsmeister 2015: 61). Ogni marca può poi contenere ulteriori specificazioni attraverso l'uso di attributi: se, per esempio, si vuole mettere in evidenza che la frase nell'esempio (2) è una citazione in lingua italiana, può essere usata la marca *q* (*q* = citazione), specificata dall'attributo *lang*, che sta per "lingua", e dal suo contenuto *it*, che sta per "italiano":

(2) <q lang='it'>una raccolta strutturata di testi in formato elettronico che si assumono rappresentativi di una data lingua o di un suo sottoinsieme, mirata ad analisi di tipo linguistico </q>

Nel corso degli anni sono stati elaborati linguaggi di marcatura da vari enti (università, enti di ricerca, società commerciali), sviluppati anche insieme a software per il loro utilizzo. Solo negli anni Ottanta del ventesimo secolo è stato creato uno standard per la codifica informatizzata dei dati testuali: la *Text Encoding Initiative* (TEI) (cfr. Spina 2001: 80). La versione integrale della TEI comprende oltre 400 marche diverse, che permettono di descrivere ogni tipo di testo in formato elettronico. Esiste una versione ridotta chiamata *TEI Lite*, che contiene solo gli elementi basilari ed essenziali per descrivere un testo: le sue caratteristiche fisiche (parti evidenziate con il corsivo, grassetto ecc.; le interruzioni di pagina ecc.), strutturali (intestazioni, corpo del testo, titoli ecc.), logiche (citazioni, liste ecc.) e linguistiche (nomi propri, parole straniere, termini tecnici ecc.) (cfr. Spina 2001: 84s.).

Nel secondo caso, l'annotazione in senso stretto (chiamata anche etichettatura o taggatura), si concentra invece solo su informazioni di tipo linguistico-interpretativo, in un certo senso più soggettivo (cfr. Freddi 2019: 19). Non si tratta di una procedura obbligatoria in un corpus, soprattutto se l'obiettivo è quello di lavorare con la totalità dei dati puri, ossia senza che essi siano corredate da informazioni di tipo linguistico (cfr. il contestualismo). Può però essere utile nel caso in cui si vogliono ricercare fenomeni specifici o per disambiguare forme omografe. L'annotazione ha innumerevoli potenzialità, soprattutto per la redazione di grammatiche di riferimento<sup>19</sup>. Inoltre, questa procedura permette non solo la verifica dei risultati dell'analisi, ma anche la loro riproducibilità (cfr. Lemnitzer/Zinsmeister 2015: 58-60).

L'annotazione linguistica può essere fatta a livelli linguistici diversi (si veda la Tabella 1.1):

19 Un esempio è la *Longman Grammar of Spoken and Written English* (Biber et al. 1999). Cfr. Freddi (2019: 20).

Livello linguistico	Annotazione
morfologico	forma base ( <i>Lemma</i> ), forma flessa
morfosintattico	parti del discorso ( <i>Wortart</i> )
sintattico	costituenti sintattici, relazioni di dipendenza, campi topologici
semantico	nomi propri, campi semantici, metafore, <i>frames</i>
pragmatico	coreferenze, strutture informative, relazioni discorsive, connotazione ( <i>sentiment</i> )
altri	struttura del testo, ortografia, errori, aspetti fonetici e prosodici, pause, elementi paraverbali (gestualità, mimica), analisi del discorso

**Tabella 1.1:** Annotazione per livello linguistico

Il procedimento dell'annotazione presuppone una segmentazione del testo, detta anche tokenizzazione (*Tokenisierung*) in parole o token, che permette di distinguere ogni occorrenza di ciascuna forma.

La tokenizzazione viene fatta in maniera automatica dalla macchina, per la quale le parole sono sequenze ininterrotte di caratteri precedute e seguite da uno spazio bianco o da un segno di interpunzione (cfr. Freddi 2019: 20). Non sempre però la parola a livello informatico coincide con la parola a livello linguistico. Casi problematici sono, tra gli altri:

- le forme contratte costituite da preposizione + articolo determinativo, come *am* (forma contratta di *an + dem*), *ins* (forma contratta di *in + das*), *aufs* (forma contratta di *auf + das*) ecc.;
- le forme come *denkense* (= *denken Sie*), le espressioni multiparola (*Mehrwortlexemen*) come *Rote Bete* o *Gelbe Karte* dove le due unità divise dallo spazio bianco formano un'unità dal punto di vista del significato;
- le frasi idiomatiche come *Tomaten auf den Augen haben* (in italiano 'avere gli occhi foderati di prosciutto') ecc.

Anche i segni di interpunzione stessi possono essere ambigui: per esempio il punto in tedesco può avere anche altre funzioni, come si evince dai seguenti esempi: *Prof.* come abbreviazione di *Professor* e *2. April* (cfr. Lemnitzer/Zinsmeister 2015: 62). Al momento dell'etichettatura dovranno essere quindi effettuate delle scelte da esplicitare agli utenti.

Il primo tipo di annotazione (cfr. Tabella 1.1) è quella morfologica (*morphologische Annotation*), basata su un processo di codifica chiamato lemmatizzazione (*Lemmatisierung*) (cfr. Freddi 2019: 21). Con questo procedimento le varianti morfologiche di una parola vengono associate alla sua forma base, il lemma (*Lemma*), come accade di consueto nella prassi lessicografica. Per esempio, le forme della flessione verbale *gebe, gebst, geht, geben* sono da ricondursi alla forma base *geben*.

L'annotazione morfosintattica (*morphosyntaktische Annotation*), chiamata comunemente etichettatura grammaticale (in inglese *Grammatical Tagging, Part-of-speech tagging, POS tagging*), invece è quella più utilizzata. La lista di tutti i *tag* utilizzati

in un corpus si chiama *tagset*, e di norma contiene tra i 50 e 150 *tag*<sup>20</sup>, descritti usualmente nelle Indicazioni per l'Annotazione<sup>21</sup> (*Annotationsrichtlinien*) (cfr. Lemnitzer/Zinsmeister 2015: 63). Con l'annotazione morfosintattica, a ogni parola nel corpus può essere attribuita una parte del discorso: nome (*Nomen*), verbo (*Verb*), aggettivo (*Adjektiv*), avverbio (*Adverb*) ecc. La parte del discorso viene indicata nel *tag* che la delimita: per esempio, ADJ.\* sta per *aggettivo* (cfr. Freddi 2019: 21). Si possono però annotare anche altre informazioni, come:

- la distribuzione, ossia la posizione (preposizione vs. postposizione);
- la funzione sintattica (attributiva vs. predicativa);
- le caratteristiche morfologiche (verbo finito vs. verbo non finito);
- le caratteristiche semantiche (nome comune vs. nome proprio).

Per alcune parti del discorso vi sono diverse possibilità di annotazione: per esempio, la funzione attributiva e predicativa dell'aggettivo non sono solo collegate alla sua posizione all'interno della frase, ma si distinguono anche per aspetti morfologici (funzione attributiva = aggettivo flesso; funzione predicativa = aggettivo non flesso).

Di seguito (Tabella 1.2) alcuni esempi tratti dallo *Stuttgart-Tübingen Tagset* (STTS)<sup>22</sup>:

Classe di parola ( <i>Wortart</i> )	Descrizione ( <i>Beschreibung</i> )	Esempi ( <i>Beispiele</i> )
ADJA	attributives Adjektiv	[das] große [Haus]
ADJD	adverbiales oder prädikatives Adjektiv	[er fährt] schnell, [er ist] schnell
ADV	Adverb	schon, bald, doch
APPR	Präposition; Zirkumposition links	in [der Stadt], ohne [mich]
APPO	Postposition	[ihm] zufolge, [der Sache] wegen
APZR	Zirkumposition rechts	[von jetzt] an
ART	bestimmter oder unbestimmter Artikel	der, die, das, ein, eine

**Tabella 1.2:** Esempificazione di alcuni *tag* del STTS con relativa descrizione ed esempi

Con l'annotazione morfosintattica è possibile disambiguare forme omografe (come per esempio *natürlich*, che può fungere sia da aggettivo sia da avverbio), ma non sempre la scelta del *tag* è univoca. Per ovviare a ciò, alcuni dei *tag* considerati “problematici” non sono presenti nei *tagset* più comuni; è il caso,

20 In Germania si è affermato lo *Stuttgart-Tübingen Tagset* (STTS). Questo piccolo *tagset* comprende 54 *tag*.

21 Nelle Indicazioni per l'Annotazione vi sono definizioni, esempi nonché soluzioni di possibili casi problematici (cfr. Lemnitzer/Zinsmeister 2015: 61).

22 Cfr. <https://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/germantagsets/> [20.10.22].

per esempio, del *tag* usato per la distinzione tra l'aggettivo usato in funzione predicativa e l'aggettivo usato in funzione avverbiale, per il quale l'annotazione automatica potrebbe dare risultati inesatti, essendo necessaria l'analisi di tutta la frase per la disambiguazione. Le categorie della morfologia flessiva abitualmente presenti nella maggior parte dei set sono caso (*Kasus*), genere (*Genus*), numero (*Numerus*), persona (*Person*), tempo (*Tempus*) e modo (*Modus*) (cfr. Lemnitzer/Zinsmeister 2015: 68).

Un altro tipo di annotazione è quella sintattica (*parsing*). Per mettere in evidenza le relazioni tra le parti del discorso, l'annotazione sintattica può fare ricorso a strutture ad albero (*Baumbanken*) o a parentesi con etichette (*Klammerstrukturen*) (cfr. Freddi 2019: 21-22). Con l'annotazione sintattica si possono annotare sia i costituenti sintattici (per esempio il sintagma nominale, verbale ecc.) e le loro funzioni grammaticali (per esempio soggetto, oggetto diretto, oggetto indiretto) – in questo caso si parla di *Konstituentenstruktur*<sup>23</sup> – sia le relazioni di dipendenza gerarchica tra il verbo (predicato) e i suoi argomenti (*Dependenzstruktur*)<sup>24</sup>. Uno sviluppo più recente di questo tipo di annotazione è il *Chunking* (*Partial Parsing*), in cui si analizzano porzioni di frase<sup>25</sup>.

Altra tipologia di annotazione è quella semantica, utilizzata soprattutto per la descrizione del lessico in caso di realizzazione di tesauri organizzati per campi semantici, per studi sul lessico specialistico o sulle metafore (cfr. Freddi 2019: 23). I singoli significati delle parole (*Lesarten*) si possono annotare sulle parti del discorso, così che una parola possa essere associata a vari significati, come in una ontologia o in un dizionario<sup>26</sup>. L'annotazione di *frames*<sup>27</sup> (cfr. la *Frame*

23 Corpora con annotazione dei costituenti sono la TIGER-*Baumbank* (<https://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger/>) e le tre TüBa-*Baumbanken* (TüBa-D/Z, TüBa-D/DP e TüBa-D/W). Cfr. <https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/seminar-fuer-sprachwissenschaft/arbeitsbereiche/allg-sprachwissenschaft-computerlinguistik/ressourcen/korpora/> [20.10.22].

24 Corpora con annotazione delle relazioni di dipendenza sono la *Prague Dependency Treebank* (cfr. <https://ufal.mff.cuni.cz/pdt2.0>) [20.10.22], in cui ogni verbo è analizzato nelle sue relazioni gerarchiche con le altre parole della frase, annotate seguendo la tradizione sintattica della scuola di Praga. Un corpus in ambito germanofono è la *Hamburger Dependency Treebank* (cfr. Foth et al. 2014).

25 Un corpus con annotazione di *chunks* è il *Tübinger Partiiell Geparste Korpus des Deutschen/Schriftsprache*.

26 Un esempio è il corpus *OntoNotes*, in cui le parti del discorso sono annotate semanticamente in testi inglesi e cinesi con link a *WordNet* (<https://wordnet.princeton.edu>) [20.10.22].

27 Cfr. *FrameNet* (<https://framenet.icsi.berkeley.edu>) [20.10.22], un database lessicale, ossia un indice di unità lessicali e dei loro pattern sintattici. Il progetto mira a individuare pattern associativi tra significati e relazioni sintattiche, con lo scopo di compilare un inventario di *frames* semantici (cfr. Freddi 2019: 39). In ambito germanofono si è sviluppato il *German FrameNet*, diramatosi in tre gruppi: il progetto SALSa di Saarbrücken (<https://www.coli.uni-saarland.de/projects/salsa/> [20.10.22]), il gruppo di lavoro a Stoccarda intorno a U. Heid con focus su collocazioni e nominalizzazioni e il gruppo di Austin, diretto da H.C. Boas che utilizza i dati di Saarbrücken per creare un database sul modello di *FrameNet*.

*Semantik* di Fillmore 1968), invece, va oltre il livello della parola (cfr. Lemnitzer/Zinsmeister 2015: 81).

L'annotazione pragmatica, con la quale si oltrepassano non solo i confini della parola ma anche della frase, si concentra soprattutto sugli atti linguistici in riferimento agli enunciati (assertivi, direttivi, commissivi ecc.), svolgendo un ruolo importante per l'identificazione di pattern argomentativi (cfr. Freddi 2019: 23). Questo tipo di annotazione riguarda anche le anafore, ossia le relazioni di coreferenza: una anafora, per esempio un pronome o un sintagma nominale, viene messa in relazione con il suo antecedente (*Bezugswort*, *Antezedens*). In un corpus possono essere annotate le relazioni di coreferenza<sup>28</sup> per analizzare, per esempio, le catene di referenti (*Referenzketten*): quali espressioni vengono usate per identificare persone, oggetti, eventi affinché il testo sia interpretato coerentemente? (cfr. Lemnitzer/Zinsmeister 2015: 83). Un'ulteriore possibilità di annotazione riguarda la struttura informativa, in cui si focalizzano sia il tema della frase o dell'espressione, sia la nuova informazione (cfr. il *Potsdam Commentary Corpus*)<sup>29</sup>.

In passato l'annotazione era molto dispendiosa, non solo in termini economici, ma anche in termini di tempo, essendo prevalentemente manuale, ma oggi esistono strumenti automatici e semiautomatici che facilitano questo tipo di procedura. Il controllo manuale<sup>30</sup> di annotazioni semiautomatiche può avvenire sia in maniera sincrona (durante l'annotazione, in caso di programmi interattivi), sia in maniera asincrona, in una fase successiva. Vi è anche la possibilità che non vi sia alcun tipo di annotazione manuale e che i dati vengano usati direttamente dopo l'annotazione automatica (si veda l'annotazione automatica di software come *Sketch Engine*<sup>31</sup>, *AntConc*<sup>32</sup>). L'annotazione morfosintattica, per esempio, può raggiungere un grado di accuratezza tra il 95% e il 98% (cfr. Lemnitzer/Zinsmeister 2015: 68), dato che gli strumenti sono in grado di disambiguare le parti del discorso. Gli altri tipi di annotazione hanno invece una precisione più bassa e richiedono maggior revisione manuale. Più si procede verso analisi maggiormente astratte, più l'annotazione diventa inevitabilmente manuale (cfr. Freddi 2019: 19).

28 Per l'annotazione della coreferenza si veda Poesio (2004). Un corpus in cui è annotata la coreferenza è il corpus DIRNDL (cfr. Björkelund et al. 2014), costituito da telegiornali radio letti (sono presenti dati audio e trascrizioni): <https://www.ims.uni-stuttgart.de/en/research/resources/corpora/dirndl/> [03.12.2023].

29 Cfr. <http://angcl.ling.uni-potsdam.de/resources/pcc.html> [21.10.2022].

30 Anche l'annotazione manuale non è priva di errori. Se, per esempio, più annotatori annotano i medesimi dati, non vi è mai una corrispondenza del 100%. Per questo, oltre al dato percentuale, viene dato molto spesso come valore di riferimento (ossia l'*Inter-Annotator Agreement*) il valore  $k$  ("kappa"), che tiene conto del fatto che una parte delle corrispondenze è dovuta a fattori casuali (cfr. Perkuhn/Keibel/Kupietz 2012).

31 Cfr. <https://app.sketchengine.eu/> [21.10.2022].

32 Cfr. <https://www.laurenceanthony.net/software/antconc/> [21.10.2022].

### 1.3 Approcci di analisi di un corpus

Nella linguistica dei corpora si possono distinguere diversi tipi di approcci per l'analisi di un corpus. Elena Tognini-Bonelli (2001) opera la nota distinzione tra *corpus-based*, in cui i dati vengono consultati per verificare ipotesi o cercare esempi, e *corpus-driven*, dove i dati vengono usati nella loro interezza e le conclusioni sono tratte esclusivamente da essi:

The corpus, therefore, is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well defined system. The theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus. (Tognini-Bonelli 2001: 84)

Il termine inglese *corpus-based* non deve essere però confuso con il termine tedesco *korpusbasiert*<sup>33</sup>, che è invece l'equivalente di *corpus-driven*, ragion per cui Lemnitzer/Zinsmeister (2015: 34) propongono la distinzione tra *korpusbasiert quantitativ* (1), *korpusbasiert quantitativ und qualitativ* (2) e *korpusgestützt* (3):

(1) L'approccio *korpusbasiert quantitativ* utilizza corpora grezzi (ossia non annotati) e opera per mezzo di modelli statistici. I dati vengono estratti, ma non interpretati. Analisi tipiche sono l'identificazione della frequenza assoluta e relativa di unità linguistiche in un corpus, l'estrazione di *keywords* singole e multiple di un corpus, l'estrazione della distribuzione di unità linguistiche, l'estrazione di *n-grams*<sup>34</sup>, l'estrazione di parole simili da un punto di vista semantico in base alla co-occorrenza. Questo approccio trova applicazione soprattutto nell'*information retrieval*;

(2) L'approccio *korpusbasiert quantitativ und qualitativ* utilizza corpora grezzi (ossia non annotati) e si concentra anche su aspetti semantici. I dati estratti vengono necessariamente interpretati. Tutte le procedure dell'approccio (1) possono valere anche per l'approccio (2), ma ne segue una interpretazione. Le analisi tipiche si concentrano anche sulla co-occorrenza, sulle collocazioni e sui campi semantici. Questo approccio trova applicazione in lessicografia, nella didattica delle lingue straniere e nelle scienze della traduzione (cfr. il contestualismo);

(3) L'approccio *korpusgestützt* fa uso di corpora annotati o di raccolte di esempi (*Belegsammlungen*); si concentra prevalentemente su aspetti sintattici e viene utilizzato per verificare o confermare ipotesi, partendo quindi da aspetti teorici. Un'analisi tipica si concentra sulla posizione delle parti del discorso in una frase. Questo approccio trova applicazione soprattutto nella linguistica teorica (cfr. il generativismo).

33 Altro errore molto comune è l'uso della parola tedesca *korpusbasiert* per tradurre l'italiano 'basato sui corpora', il cui equivalente sarebbe in realtà *korpusgestützt*.

34 Per *n-grams* si intendono una serie N di parole che si succedono una dopo l'altra nel medesimo ordine, dove N sta per un qualsiasi numero > di 0. Si parla di *Bigrams* (2-Grams), di *Trigrams* (3-Grams) eccetera. Cfr. Manning/Schütze (2002).



# Capitolo 2

## Corpora e analisi linguistica

In questo capitolo sono presentati alcuni concetti fondamentali della linguistica dei corpora necessari per affrontare analisi linguistiche sia di tipo *corpus based*, sia di tipo *corpus driven* (cfr. paragrafo 1.3). In primo luogo, si procede a una distinzione tra token, type e lemma (paragrafo 2.1); successivamente viene presentato il procedimento di ricerca in un corpus per poi introdurre le espressioni regolari, utili per ricerche più complesse (paragrafo 2.2). Nel paragrafo 2.3 il focus verte sui possibili formati di visualizzazione dei risultati di una ricerca (KWIC e testo) e sulle possibilità di esportazione dei dati. Il paragrafo 2.4 si occupa, invece, dell'estrazione delle liste di frequenza (parole, lemmi ecc.). Il paragrafo 2.5 si concentra sulle liste di parole chiave (*keywords* singole e multiple) e sugli *n-grams*, mentre il paragrafo 2.6 sul profilo di parola. Il capitolo si conclude con il paragrafo 2.7 in cui si prendono in esame ulteriori possibilità di analisi, come il confronto tra i profili di parola tramite *Word Sketch Difference* e l'estrazione di liste di parole simili dal punto di vista semantico con *Thesaurus*.

### 2.1 Token, type e lemmi

Per poter svolgere una ricerca empirica in un corpus è necessario conoscere le unità che lo compongono.

Nel paragrafo 1.2 sono già stati introdotti i dati primari, ossia i testi scritti e le espressioni orali, o le loro porzioni, che costituiscono il corpus. I dati primari vengono poi segmentati automaticamente<sup>1</sup> dai sistemi informatici in unità sempre più piccole: in paragrafi (*Absätze*), in frasi (*Sätze*) e in parole<sup>2</sup> (*Wörter*).

Ai fini dell'identificazione delle parole (*Wörter*) all'interno della frase, i principali software applicano usualmente la seguente regola<sup>3</sup>: le parole sono sequenze

---

1 La segmentazione automatica ha raggiunto oggi risultati sempre più precisi; tuttavia, essendo probabile un certo margine di errore, è necessario conoscere quali criteri sono stati utilizzati.

2 Lemnitzer/Zinsmeister (2015: 35) preferiscono usare il termine “sequenza di lettere” (*Zeichenkette*) rispetto a “parola” (*Wort*). Le molteplici definizioni di parola che si trovano in letteratura (cfr. anche paragrafo 1.1.1 della parte II) e che si soffermano su aspetti semantici, morfosintattici, sintattici (cfr. tra gli altri Glinz 1973, Bloomfield 1933, Lyons 1968 ecc.) non sono, infatti, sempre adeguate all'accezione usata nella linguistica dei corpora. La definizione che vi si avvicina di più è quella grafematica di Bühler et al. (1972), per i quali «Wörter sind sprachliche Einheiten, die als Folge von Buchstaben zwischen zwei Spatien geschrieben werden» (Perkuhn/Keibel/Kupietz 2012: 26).

3 Oltre alle riflessioni che seguiranno, occorre considerare che alcuni sistemi di analisi contano anche i segni di interpunzione e i simboli come token. È opportuno quindi informarsi sui criteri adottati dal sistema per la tokenizzazione.

di lettere precedute e seguite da uno spazio. Questo assunto, applicabile al tedesco, potrà subire lievi variazioni in altre lingue; nel caso dell'italiano, infatti, la cui struttura è più complessa di altre lingue<sup>4</sup>, vale di norma la seguente regola: la parola è «una sequenza di parole preceduta e seguita da uno spazio e/o da un apostrofo» (Spina 2001: 108). Prendendo quindi come esempio *all'amatriciana*, si contano due parole: *all* e *amatriciana*.

Una volta stabilita la regola per distinguere le parole all'interno di una frase, vi sono però altre considerazioni da fare. Si osservi, per esempio, la seguente frase (1), contando le parole secondo l'indicazione di cui sopra:

(1) Eine leckere Torte ist eine leckere Torte, zwei Torten sind noch leckerer<sup>5</sup>.

Le possibilità sono molteplici (cfr. Tabella 2.1):

1. si contano tutte le parole, ossia tutte le sequenze precedute e seguite da uno spazio una sola volta, anche se esattamente identiche tra loro e in posizione diversa. In questo caso sono 12 parole (riga a);
2. si contano tutte le parole uguali una volta sola, anche se si trovano in posizione diversa. In questo caso sono 10 (riga b);
3. si contano le parole facendo riferimento a un'entità astratta. In questo caso sono 6 (riga c):

	Esemplificazione	N.	Tipo di parola
a	Eine, leckere, Torte, ist, eine, leckere, Torte, zwei, Torten, sind, noch, leckerer	12	token ( <i>Token</i> )
b	Eine, leckere (2x), Torte (2x), ist, eine, zwei, Torten, sind, noch, leckerer	10	type ( <i>Type</i> )
c	eine (Eine, eine), leckere (2x leckere, leckerer), Torte (2x Torte, Torten), sein (ist, sind), zwei, noch	6	lemma ( <i>Lexem-Type</i> )

**Tabella 2.1:** Conteggio delle parole dell'esempio (1) in base ai criteri 1., 2. e 3.

Nel primo caso (riga a) le parole sono state contate secondo la definizione grafematica; è possibile affermare che corrispondono ai token. Nel secondo caso (riga b) le parole sono state contate in base alla loro forma, indipendentemente dalla loro ripetizione e dalla loro posizione: token uguali vengono ascritti alla medesima entità, ossia al type. A volte però si può fare riferimento a un'entità ancora più astratta, ossia il lemma (riga c), dove per lemma si intende la forma base del type.

4 Spina (2001: 108) sottolinea il fatto che, se un corpus non è annotato e lemmatizzato, la sola tokenizzazione porterebbe a conteggiare parole come *sulla* o *portaci* come singole occorrenze.

5 L'esempio (1) è stato liberamente ispirato da Perkuhn/Keibel/Kupietz (2012: 27).

Token, type e lemma sono categorie di analisi centrali per lo studio delle frequenze. Se si osservano le posizioni più alte delle liste di frequenza si trovano abitualmente parole vuote (*Synsemantika*), come articoli, congiunzioni ecc. Solo in corpora piccoli si possono trovare nelle prime venti posizioni parole portatrici di significato (*Autosemantika*)<sup>6</sup>; nei corpora più estesi le parole portatrici di significato si trovano, invece, a partire dal quarantesimo o cinquantesimo posto della lista. Maggiore è la quantità di dati, maggiore è, infatti, la tendenza delle parole a frequenza più alta a stabilizzarsi, ossia a occorrere nelle medesime posizioni della lista (cfr. Spina 2001: 109)<sup>7</sup>.

Si osservino, a titolo esemplificativo, le prime venti posizioni delle seguenti liste di frequenze estratte da due corpora di lingua tedesca (tool *Sketch Engine*), *German Web 2020* e *EUR-Lex German 2/2016* (Tabella 2.2):

N.	German Web 2020	EUR-Lex German 2/2016
1	,	,
2	.	der
3	die	die
4	der	.
5	und	)
6	in	und
7	“	(
8	das	in
9	den	des
10	mit	/
11	zu	von
12	von	für
13	ist	den
14	:	zu
15	für	:
16	auf	auf
17	im	oder
18	nicht	“
19	ein	I
20	)	im

**Tabella 2.2:** Le 20 parole più frequenti in *German Web 2020* e in *EUR-Lex German 2/2016*

<sup>6</sup> Cfr. la *Dudengrammatik* (cfr. Duden 2016: 581).

<sup>7</sup> Secondo la legge di Zipf il rapporto tra rango e frequenza in corpus è regolare (cfr. Zipf 1935; 1949).

Anche se si tratta di due corpora molto diversi tra loro (*German Web 2020* è un corpus estratto automaticamente da Internet, mentre *EUR-Lex German 2/2016* un corpus specialistico), si può notare che molte forme si ritrovano in entrambi i corpora. Dalla Tabella 2.2 emerge, inoltre, che anche i segni di interpunzione, le parentesi e le virgolette vengono considerati token dal sistema.

Se si andassero ad analizzare corpora ancora più grandi, strutturati, lemmatizzati e annotati dal punto di vista morfosintattico, si otterrebbero i medesimi risultati (cfr. Spina 2001: 109). Proprio questa base costante può essere un punto di partenza per l'analisi dei risultati, che, se si discostano, possono essere indice di anomalie.

Un altro aspetto da tenere in considerazione è il confronto tra i dati. I valori di frequenza, infatti, non devono essere considerati in maniera isolata, ma messi a confronto. Il dato assoluto non è sempre un'informazione rilevante: vedere che una parola X occorre un determinato numero di volte in un corpus Y non è spesso sufficiente. L'occorrenza di una parola, infatti, deve essere messa in rapporto con la frequenza delle altre parole del corpus: la parola X corrisponde cioè a un determinato numero di occorrenze, che rappresenta una certa percentuale dei dati. Spina (2001: 114) riflette, ad esempio, sul fatto che i type di un corpus con una certa estensione sembrano essere costituiti da tre gruppi: un numero molto ristretto di parole vuote molto frequenti (ca. il 40% del corpus); un numero molto elevato di parole con frequenza molto bassa, di solito portatrici di significato autonomo, che si trovano nei ranghi più bassi; infine, un numero di parole né eccessivamente numerose, né eccessivamente frequenti, che rappresentano una minoranza.

Anche il rapporto tra un determinato type e i token in un corpus è un dato su cui porre l'attenzione. Per esempio, se *Haus* occorre X volte in un corpus composto da un numero Y di token, il rapporto si calcola nel seguente modo:

$$\text{Type/token} = X (\text{frequenza del type}) / Y (\text{numero totale di token del corpus})$$

Per ottenere la percentuale occorre moltiplicare il numero ottenuto per 100. Supponendo che *Haus* occorra 150 volte in un corpus di 60.000 token, allora il rapporto type/token è 0,0025 e la percentuale è lo 0,25% (0,0025 x 100). Questo rapporto è calcolato automaticamente da molti software di analisi.

Anche il modo in cui i type si distribuiscono all'interno dei token è un parametro rilevante. Infatti, dal rapporto tra il numero totale di tutte occorrenze (token) e il numero totale di tutte forme distinte in un corpus (il numero di type), ossia la *type token ratio* (TTR), è possibile fare considerazioni sulla varietà lessicale del corpus (cfr. Freddi 2019: 89). Più il rapporto è basso, meno ricco è il vocabolario di un corpus (ci sono pochi type che si ripetono un numero elevato di volte); più il rapporto è grande, maggiore è la differenziazione lessicale e minore è la ripetitività (cfr. Spina 2001: 116). La TTR rappresenta dunque

un'ottima possibilità per esplorare la variazione stilistica (cfr. Freddi 2019: 90). I linguaggi settoriali saranno caratterizzati da pochissima variabilità, così come un singolo genere testuale.

Il calcolo della frequenza degli elementi che compongono un corpus è utile anche per identificare il grado di leggibilità, ossia di comprensibilità dei suoi dati primari (cfr. Spina 2001: 120).

Oltre alla frequenza devono essere considerate anche la rilevanza di una attestazione (*Relevanz*) e la sua distribuzione (*Streuung*), ad esempio in sottocorpora.

Nonostante le numerose possibilità di riflessione offerte da questo tipo di analisi, occorre considerarne anche i limiti. I dati numerici, infatti, non forniscono informazioni sul fatto che l'unità possa essere parte di strutture più ampie (si pensi ai verbi separabili) e non si distinguono, al loro interno, le differenze di significato e di categoria grammaticale: per esempio, in tedesco la parola *sein* può essere sia la forma infinita del verbo *essere*, sia l'aggettivo possessivo di terza persona singolare. Inoltre, a partire dai soli dati numerici, non è possibile riflettere sul contesto d'uso della parola; infine, la ricerca di combinazioni lessicali<sup>8</sup>, in alcuni sistemi, può rivelarsi problematica.

In ogni caso, i dati numerici rappresentano un buon punto di partenza, nonché un filtro preliminare per selezione di aspetti da approfondire.

## 2.2 Il procedimento di ricerca in un corpus

La ricerca in un corpus può avvenire seguendo diverse modalità ed è connessa al sistema di gestione e analisi<sup>9</sup> (*Korpusrecherchesystem*) utilizzato.

Alcune modalità di ricerca sono possibili in molti sistemi, come per esempio:

1. la ricerca di una forma di parola: nel corpus si cerca esattamente la forma dell'unità linguistica d'interesse e la si digita nella maschera di ricerca. Digitando *Hunde*, il sistema cerca quindi tutte le occorrenze di *Hunde* (non di *Hund* però). Nel tool *Sketch Engine* questa opzione corrisponde alla ricerca per *word* nella funzione *Concordance*; in COSMAS II la ricerca per forma di parola è impostata di default; digitando *Hunde*, in DWDS la ricerca viene dirottata su *Hund*;

8 Fino all'introduzione del concetto di collocazione (cfr. paragrafo 2.6) e delle sue varie possibilità di definizione e interpretazione, si utilizza il termine generico 'combinazione lessicale'.

9 Oggi i sistemi di analisi di corpora utilizzabili sono molteplici. Sebbene nel presente volume si farà principalmente riferimento a *Sketch Engine* ([www.sketchengine.eu](http://www.sketchengine.eu)), che può essere usato online, occorre menzionare anche *AntConc* di Laurence Anthony ([www.laurenceanthony.net/software/antconc/](http://www.laurenceanthony.net/software/antconc/)), *WordSmith* di Mike Scott (<https://www.lexically.net/wordsmith/>), *Corpus Explorer* di Jan Oliver Rüdiger (<https://notes.jan-oliver-ruediger.de/software/corpus-explorer-overview/>). Quest'ultimi possono essere installati sul proprio computer. Mentre *AntConc* e *Corpus Explorer* sono gratuiti, *Sketch Engine* e *WordSmith* sono a pagamento (per entrambi vi sono versioni *free source* con meno funzionalità, ma interessanti per fare dei test). Tutti e tre i sistemi sono adatti all'analisi di concordanze. Le analisi sintattiche sono invece possibili tramite *TIGER Search*, *TünDRA* e *ANNIS* (cfr. Lemnitzer/Zinsmeister 2015: 94).

2. la ricerca di un lemma: nel corpus si ricerca la forma base. Se nella maschera di ricerca si digita *Hund*, il sistema mostra anche *Hunde*, *Hundes*, *Hunden* eccetera. Nel tool *Sketch Engine* questa opzione corrisponde alla ricerca per *lemma* nella funzione *Concordance*; in COSMAS II non è possibile farla senza l'uso di espressioni regolari (*reguläre Ausdrücke*), e in questo caso la ricerca sarà *&Hund*; in DWDS questa ricerca è impostata di default;
3. la ricerca di una sequenza di parole: nel tool *Sketch Engine* questa opzione corrisponde alla ricerca per *phrase* nella funzione *Concordance*: ad esempio, digitando *bellender Hund* verranno mostrate tutte occorrenze di *bellender + Hund*, dove vi è variazione del sostantivo *Hund/Hunde*; digitando *Zähne putzen* il sistema cerca tutte le occorrenze di *Zähne putzen* ([...] *Zähne geputzt*, [...] *Zähne putze*). Tra *bellender* e *Hund*, come pure tra *Zähne* e *putzen*, non vi sarà mai altro materiale lessicale. Questo tipo di ricerca si adatta quindi solo per un certo tipo di combinazioni lessicali. In COSMAS II non è possibile farla senza l'uso di espressioni regolari; nel DWDS è possibile solo la ricerca di combinazioni lessicali fisse come *Zähne putzen*, oppure *Hunde, die bellen, beißen nicht*;
4. la ricerca di una sequenza di lettere: nel tool *Sketch Engine* questa opzione corrisponde alla ricerca per *character* nella funzione *Concordance*. Se si digita *ling* verranno mostrate tutte le forme di *ling* (se ovviamente presenti nella lingua del corpus), e tutte le parole che contengono *ling*, indipendentemente dalla posizione della sequenza nella parola (*Linguistik, Lieblingsessen, Bowling* ecc.); in COSMAS II si possono solo cercare le forme di *ling*, ma non la sequenza all'interno di una parola senza l'uso di espressioni regolari; in DWDS la ricerca non conduce a nessun risultato. Se, invece, si volesse cercare il suffisso *-ling* in *Sketch Engine* e in COSMAS II, si dovrebbero usare le espressioni regolari: *\*ling*; in DWDS il suffisso si potrebbe cercare con l'opzione *-ling*;
5. la ricerca di una parola, di una sequenza di parole con indicazione della loro relazione sintattica con l'uso della *corpus query language* (CQL). Ogni sistema mette a disposizione indicazioni sulla sua sintassi per la formulazione della ricerca. In *Sketch Engine* può essere fatta scegliendo l'opzione *CQL* nella funzione *Concordance*; in COSMAS II e in DWDS si digita la stringa di ricerca nella maschera di ricerca (per entrambi i corpora ci sono indicazioni nella *homepage*). In questo tipo di ricerca si può fare uso sia di *tag* (cfr. paragrafo 1.2), sia di espressioni regolari.

Come evidenziato, le espressioni regolari sono indispensabili per alcune modalità di ricerca. Esse si distinguono usualmente in simboli (*Platzhalterzeichen, wildcard*) e operatori (*Operatoren*) e il loro uso oggi è permesso in quasi tutti i sistemi<sup>10</sup>. Esse possono svolgere varie funzioni, permettendo di:

10 Anche se presenti, le differenze tra un sistema e l'altro sono oggi minime. Per Perkuhn/Keibel/Kupietz (2012: 36) le espressioni regolari nel linguaggio informatico *Perl* sono considerate quasi uno standard.

- cercare per esempio due alternative (dem | den; Rat | Rad);
- indicare qualcosa come opzionale;
- indicare che qualcosa viene ripetuto;
- inserire uno slot da riempire (*Platzhalter*)<sup>11</sup>: per esempio con Ra. si estraggono *Rad, Rap, Rat* eccetera;
- filtrare tutte le parole che iniziano con una sequenza un\*, che finiscono con una certa sequenza \*ling, o che al loro interno hanno una certa sequenza \*mitte\*;
- indicare che tra una parola e un'altra ci deve essere una certa distanza (un certo numero di parole);
- escludere la presenza di certe lettere, parole.

Le espressioni regolari permettono quindi di cercare in un corpus non solo parole specifiche, ma anche modelli (*Muster*), e sono indispensabili per molte ricerche.

## 2.3 Risultati della ricerca

Premesso che nella linguistica dei corpora e in analisi di tipo empirico tutti i risultati hanno il medesimo valore (cfr. Tognini-Bonelli 2001) e sono un punto di partenza per riflessioni di vario tipo, una volta avviata la ricerca tramite il motore di ricerca (*Suchwerkzeug* in tedesco, *Query Tool* in inglese) possono verificarsi due scenari:

1. la ricerca non produce alcun risultato;
2. la ricerca produce uno o più risultati.

Il primo scenario può essere causato da un errore nella formulazione della ricerca (errore di battitura, nell'uso di espressioni regolari e *tag* ecc.) o dal fatto che l'entità oggetto della ricerca non è presente nel corpus analizzato.

Nel secondo scenario i risultati della ricerca potranno essere vari. Nel caso si abbia un solo risultato, si parla di *Hapax legomena* (al singolare *Hapax legomenon*), che possono essere molto importanti per l'analisi<sup>12</sup> di neologismi e occasionalismi<sup>13</sup>; qualora si abbiano invece molti risultati, questi andranno filtrati e analizzati.

11 Per una panoramica delle espressioni regolari cfr. Perkuhn/Keibel/Kupietz (2012: 35).

12 Per riflessioni sui neologismi e gli occasionalismi per mezzo della linguistica dei corpora cfr. Lemnitzer/Zinsmeister (2015: 172-174).

13 Per occasionalismi si intendono si intendono «Wortbildungsprodukte, die nur für eine bestimmte kommunikative Gelegenheit geschaffen werden» (<https://grammis.ids-mannheim.de/terminologie/4>). Nel caso in cui l'occasionalismo poi si lessicalizzi, si parla invece di neologismo.

Il formato di visualizzazione più comune in tutti i sistemi di ricerca è la concordanza (*Konkordanz*), chiamata anche *Keyword in Context* (KWIC).

Le concordanze sono una lista (indice, elenco) di tutte le occorrenze di una parola (chiamata “parola chiave”, *keyword* in inglese, termine diffuso anche nel tedesco) nell’ambiente testuale che la circonda, chiamato cotesto (cfr. Freddi 2019: 96) o anche contesto (cfr. Spina 2001: 127)<sup>14</sup>. La *keyword*, di solito colorata (in *Sketch Engine* è rossa, si veda Figura 2.1), è allineata al centro a distanza fissa dalle porzioni che la precedono e la seguono (cfr. Freddi 2019: 96).

Insieme alle liste di frequenza (cfr. paragrafo 2.4) le concordanze sono state una delle prime forme di analisi quantitativa di corpora linguistici (cfr. Spina 2001: 127) e rappresentano oggi il cuore di ogni tipo di analisi linguistica. Sono, infatti, considerate un “s sofisticato metodo di ricerca e visualizzazione di parole chiave all’interno del loro contesto, che permette di individuare regolarità, anomalie, usi più o meno ricorrenti di determinati segmenti di lingua” (Spina 2001: 133).



**Figura 2.1:** Screenshot delle prime cinque KWIC di *Klima* nel corpus *German Web 2020* in *Sketch Engine*

Nella Figura 2.1 sono visibili le concordanze della parola *Klima* (la parola oggetto della ricerca)<sup>15</sup>. Le KWIC favoriscono la messa a fuoco dell’informazione principale presentandola in una posizione sempre uguale e centrale, e rappresentano un metodo efficace per individuare e studiare la ricorrenza delle combinazioni della parola digitata (cfr. Spina 2001: 128). Dalla KWIC nella Figura 2.1 è possibile, per esempio, fare ipotesi sul significato della parola *Klima*, che può avere varie accezioni, nonché su quali parole vengono usate preferibilmente in combinazione con essa. Si può evincere che *Klima* può accompagnarsi con aggettivi quali *politisch* e *geistig*, assumendo quindi un significato metaforico; oppure si può osservare che il sostantivo *Klima* può essere usato con il verbo *schaffen* (*ein Klima X schaffen*) ecc.

Le KWIC possono poi essere riorganizzate in ordine alfabetico sulla prima parola del contesto a sinistra (*left context*), sulla prima parola del contesto a destra

14 In linguistica i termini ‘cotesto’ e ‘contesto’ hanno significati differenti, mentre nella linguistica dei corpora identificano entrambi quella parte di testo della concordanza che si trova a sinistra e a destra della parola chiave (chiamata anche nodo). In molti sistemi di analisi di corpora questo intervallo di testo è fissato per default dal sistema (un numero X di token che precedono e seguono la parola cercata), ma può essere modificato in base agli obiettivi della ricerca.

15 La sequenza di colore blu `</s>` indica che la frase (s = *sentence*) è terminata. La sequenza blu `<s>` indica invece che la frase sta iniziando.

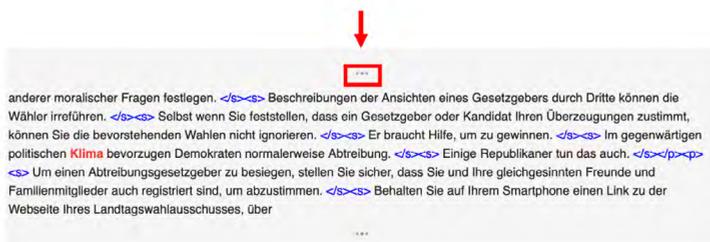
(*right context*) o sulla forma della parola centrale, la *keyword*, in base agli obiettivi di ricerca e alla lingua. Il nuovo orientamento permette di visualizzare la ricorrenza di particolari combinazioni. Per esempio, se:

- si orientano le KWIC di *Klima* secondo il contesto di sinistra (*left context*), si potrà notare che il sostantivo *Klima* può essere preceduto da aggettivi quali *ideal*, *ideologisch*, *idyllisch* (l'ordine è alfabetico);
- si orientano le KWIC secondo il contesto di destra (*right context*), si osserva, invece, che *Klima* è usato spesso nella combinazione *Klima- und Energiepolitik*, che *Klima* si combina spesso con *Informationen* (*Klima Informationen*);
- si orientano le KWIC sulla forma della *keyword*, potranno essere estrapolati i vari usi della parola *Klima* e delle sue modificazioni, come *Klimas*.

Le interrogazioni delle KWIC possono poi essere raffinate usando le funzionalità consentite dal sistema di interrogazione. L'impostazione dei criteri di ricerca riveste un ruolo molto importante, soprattutto se l'obiettivo dell'analisi non è la parola singola, ma un gruppo di forme, come nel caso dell'individuazione di un campo semantico, di una struttura sintattica eccetera (cfr. Spina 2001: 130). Inoltre, soprattutto quando si lavora con corpora molto ampi, è necessario eliminare al massimo il rumore, ossia le forme che vengono comprese in una concordanza per analogia grafica, ma che non hanno a che vedere con l'obiettivo della ricerca (cfr. Spina 2001: 128).

I risultati di una ricerca possono essere anche migliaia di KWIC; dato che, in tal caso, è difficile orientarsi, in molti sistemi (*Sketch Engine*, COSMAS II), vi è la possibilità di estrarre un campione (*sample*) di X attestazioni (per esempio 200, 500 ecc.) selezionate in maniera casuale (*get random sample*).

Oltre alle KWIC, in molti sistemi è possibile visualizzare anche una porzione di testo maggiore; in *Sketch Engine*, cliccando sulla *keyword*, si apre una finestra in cui si può visualizzare una porzione più estesa di testo<sup>16</sup> (Figura 2.2):



**Figura 2.2:** Screenshot di una porzione di testo di *Klima* nel corpus *German Web 2020* in *Sketch Engine*

<sup>16</sup> Cliccando sui tre punti in alto e in basso è possibile aumentare la porzione di testo.

In *Sketch Engine* è possibile, inoltre, impostare la visualizzazione anche per frase (*sentence*)<sup>17</sup> (Figura 2.3):



**Figura 2.3:** Screenshot delle prime cinque frasi di *Klima* nel corpus *German Web 2020* in *Sketch Engine*

Dalla Figura 2.3 sono visibili, inoltre, anche le fonti da cui l'attestazione è tratta: si nota, per esempio, che l'attestazione 1 è stata estratta dalla pagina web *xobor.de*.

I risultati della ricerca<sup>18</sup> possono poi essere salvati o esportati (ossia scaricati) in diversi formati (.txt, .csv, .xlsx, .xml ecc.) per essere analizzati empiricamente. È proprio in questa fase che i dati possono essere interpretati e utilizzati per rispondere alle proprie domande di ricerca.

## 2.4 Liste di frequenza

Come è stato già accennato nel paragrafo 2.1, il dato assoluto nell'analisi linguistica ha un valore limitato<sup>19</sup>, ma rappresenta un buon punto di partenza. È proprio la frequenza assoluta, chiamata anche *raw frequency* in inglese (*absolute Häufigkeit* in tedesco), quella che viene indicata quando si avvia una ricerca: essa indica il numero delle occorrenze di una determinata unità lessicale in un corpus.

Se si prendono come esempio le occorrenze della parola *Klima* in *German Web 2020* (20.999.598.683 token), sono 540.863 (cfr. Figura 2.4).



**Figura 2.4:** Screenshot delle frequenze di *Klima* nel corpus *German Web 2020* in *Sketch Engine*

<sup>17</sup> Per agevolare la scelta di frasi più semplici, per finalità lessicografiche e didattiche, può essere usata la funzionalità GDEX (*Good Dictionary Example*) visibile nei simboli sopra le frasi (Figura 2.3).

<sup>18</sup> In alcuni sistemi vi è una limitazione del numero di attestazioni che si possono esportare e scaricare.

<sup>19</sup> Generalizzazioni dal corpus alla lingua in toto sono possibili solo raramente (cfr. Perkuhn/Keibel/Kupietz 2012: 71).

Se poi si volesse confrontare la frequenza di una sequenza X in due corpora di dimensioni diverse, non può, però, essere presa come riferimento la frequenza assoluta, in quanto i dati non sarebbero confrontabili.

Per confrontare la frequenza di una sequenza X in due corpora di grandezze diverse, le frequenze devono essere normalizzate in base alla grandezza del corpus.

A tal fine può essere calcolata la frequenza relativa (*relative Häufigkeit* in tedesco, *relative frequency* in inglese). Il metodo più semplice è la suddivisione della frequenza assoluta per il numero di token del corpus. Se si prende come esempio la parola *Klima* in *German Web 2020* (20.999.598.683 token), la frequenza relativa è “il numero delle occorrenze di *Klima* nel corpus/il numero dei token del corpus”, ossia  $540.863/20.999.598.683 = 0.00002575587$ . Dato che il risultato di questo calcolo è composto da numeri molto piccoli, si tende a moltiplicarlo per 1.000.000 ottenendo dunque 25.75587. Questo tipo di frequenza normalizzata (*normierte Frequenz* in tedesco, *normalized frequency – nf* in inglese) viene definita “parti per milione”, abbreviata in ppm (*pmW* in tedesco). Sta a indicare che in un corpus fittizio composto da 1.000.000 token la sequenza d’interesse avrebbe la frequenza di X occorrenze. Molti sistemi come *Sketch Engine* forniscono questo dato automaticamente (Figura 2.4): la frequenza ppm di *Klima* in *German Web* è 25.76, che corrisponde a una percentuale dello 0.0026% del corpus.

Per mezzo della frequenza normalizzata si può quindi confrontare una sequenza X in due corpora diversi. Si prenda come esempio sempre la parola *Klima* in *German Web 2013*, *German Web 2018* e *German Web 2020* in *Sketch Engine* e si osservino le sue frequenze (Tabella 2.3):

	<i>German Web 2013</i> (19.808.173.163 token)	<i>German Web 2018</i> (6.382.147.542 token)	<i>German Web 2020</i> (20.999.598.683 token)
frequenza assoluta	510.649	163.501	540.863
frequenza ppm	25.78	25.62	25.76

**Tabella 2.3:** Frequenza assoluta e ppm di *Klima* nei corpora *German Web 2013*, *German Web 2018*, *German Web 2020* in *Sketch Engine*

Se ci si limitasse solo alla frequenza assoluta, si potrebbe giungere alla conclusione, errata, che la parola *Klima* sia più frequente in *German Web 2020* rispetto agli altri due corpora. In realtà, il corpus che presenta una maggiore occorrenza della parola *clima* è *German Web 2013*: questo dato si deduce proprio dalla frequenza normalizzata ppm.

Chiarita l’importanza della frequenza relativa e della frequenza normalizzata per il confronto dei dati, si può passare all’estrazione delle liste di frequenza, che in molti sistemi di gestione e analisi è automatizzata (cfr. Freddi 2019: 82). In *Sketch Engine* si può utilizzare la funzionalità *Wordlist*<sup>20</sup> per estrarre sia la lista

20 Nella configurazione attuale non si possono estrarre liste di frequenza in DeReKo e in DWDS.

delle parole più frequenti (*Wörter* in tedesco, *words* in inglese), sia la lista di frequenza lemmatizzata (*Lemmata* in tedesco, *lemmas* in inglese). Entrambe le liste sono ordinate per frequenza in ordine decrescente. In cima alla lista si trovano i valori di frequenza più alti, in fondo alla lista gli *hapax legomena*; a parità di frequenza l'ordine usuale è quello alfabetico.

Le liste di frequenza sono molto importanti per fini lessicografici e didattici: permettono, infatti, di stabilire fasce di frequenza utili per redigere lessici di frequenza di lingue o delle loro varietà (cfr. Freddi 2019: 84). Con gli studi sulle frequenze si è cercato di rispondere alla domanda: quali sono le parole indispensabili per la comprensione scritta e orale di una lingua? (cfr. Spina 2001: 120-121).

Le liste di frequenza lemmatizzate sono invece rilevanti per avere un panorama del lessico usato in un corpus e delle sue relazioni lessicali, ma ovviamente sono meno indicative delle singole forme che un lemma assume in un corpus (cfr. Freddi 2019: 85).

È inoltre possibile estrarre anche liste più specifiche, per altre categorie o fenomeni annotati, come per esempio le parti del discorso (aggettivi, preposizioni, avverbi, congiunzioni, sostantivi, numerali, verbi) in relazioni alle quali si contano le occorrenze. Se si estraessero le liste delle parti del discorso, si vedrebbe, per esempio, che le più frequenti sono l'articolo e il nome, seguiti da verbo e preposizione (cfr. Freddi 2019: 86).

Nella Figura 2.5 si può vedere uno *screenshot* della lista dei sostantivi più frequenti nel corpus *German Web 2020* in *Sketch Engine*.

	Noun	Frequency ? ↓	Frequency Per Million ? ↓
1	Jahr	38,046,562	1,811.78 ...
2	Zeit	16,736,617	797.00 ...
3	Mensch	14,644,844	697.39 ...
4	Tag	13,696,983	652.25 ...
5	Kind	12,667,525	603.23 ...
6	Frage	10,167,147	484.16 ...
7	Deutschland	9,365,968	446.01 ...
8	Seite	9,164,941	436.43 ...
9	Fall	9,111,442	433.89 ...
10	Frau	9,105,591	433.61 ...

**Figura 2.5:** *Screenshot* della lista dei sostantivi più frequenti (*wordlist, nouns*) in *German Web 2020* in *Sketch Engine*

Per ogni parola possono essere visualizzate informazioni di vario tipo, come la frequenza assoluta (*frequency*) e la frequenza ppm (*frequency per million*), necessaria per un confronto tra corpora di diversa grandezza. Cliccando poi sui tre puntini a lato di ogni riga possono essere consultate le concordanze (funzionalità *Concordance*), i profili di parola (funzionalità *Word Sketch*) (cfr. paragrafo 2.6) e le parole con profili simili (funzionalità *Thesaurus*) (cfr. paragrafo 2.7). Accedendo alle funzionalità avanzate si possono anche estrarre liste di *tag* (si estraggono i *tag* più utilizzati) o liste che focalizzano informazioni ancora più specifiche, come caso, genere, numero.

Anche se le liste di frequenza possono essere ordinate secondo modalità diverse<sup>21</sup>, l'ordinamento per frequenza assoluta, relativa o normalizzata rimane il più indicativo delle distribuzioni di frequenza delle parole nel corpus.

L'estrazione delle liste di frequenza è quindi un primo passo, che permette, tra gli altri, di isolare, all'interno di corpora molto grandi, elementi tipici o elementi di interesse (cfr. Spina 2001: 126). Successivamente si passa a un'analisi sempre più dettagliata con focus sul contesto di occorrenza, proprio perché ogni elemento di una frase è legato agli elementi che lo precedono e lo seguono. Questo procedimento dovrà essere adottato a ogni livello di analisi linguistica.

## 2.5 Liste di *keywords* e di *n-grams*

Oltre alle liste di frequenza possono essere estratte anche liste di *keywords*, (parole chiave)<sup>22</sup>, ossia le parole caratteristiche di un determinato corpus (cfr. Scott/Tribble 2006; Bondi/Scott 2010) e che occorrono con «unusual frequency», dove per «unusual» si intende «by comparison with a reference corpus of some kind» (Scott 1997: 236).

Le liste di *keywords* sono molto utili per identificare sia tematiche, sia caratteristiche lessicali e grammaticali tipiche (cfr. Stefanowitsch 2020: 354).

In *Sketch Engine* le parole chiave possono essere estratte per mezzo della funzionalità *Keywords and Terms*. Il sistema permette l'estrazione di parole chiave sia singole (*keywords*), sia multiple (*multiple keywords*); a tal fine il corpus oggetto dell'analisi (il *focus corpus*) viene confrontato con un altro corpus, considerato il corpus di riferimento (il *reference corpus*). Per default, *Sketch Engine* utilizza come corpora di riferimento i corpora appartenenti alla famiglia *TenTen* (corpora web) già presenti nel sistema. È possibile però modificare il corpus di riferimento,

21 In molti sistemi informatici le frequenze possono essere ordinate anche in senso decrescente, in ordine alfabetico e in ordine alfabetico inverso (cfr. Spina 2001: 117).

22 È opportuno precisare che «parola chiave» ha un duplice significato nella linguistica dei corpora. Può essere semplicemente la parola oggetto della ricerca e mostrata nel suo contesto (KWIC, cfr. paragrafo 2.3) oppure la parola tipica di un corpus (*keyword*), accezione utilizzata in questo paragrafo.

sostituendolo con un altro presente nel sistema o con un altro creato e inserito *ad hoc*.

Si analizzino ora le prime 10 *keywords* singole (Figura 2.6) del corpus *German Web 2020* rispetto al corpus di riferimento *German Web 2013*:

	Lemma	Frequency?		
		Focus	Reference	Score ?
1	Trump	533,400	5,862	20.4 ...
2	Coronavirus	368,953	1,768	17.0 ...
3	Corona-Krise	309,763	< 18	15.8 ...
4	erste	262,818	56	13.5 ...
5	Corona	419,892	12,558	12.8 ...
6	AfD	394,566	12,522	12.1 ...
7	Archivlink	229,431	76	11.9 ...
8	folie	260,188	2,837	11.7 ...
9	Corona-Pandemie	179,030	< 18	9.5 ...
10	DSGVO	158,555	< 18	8.5 ...

**Figura 2.6:** *Keywords* singole del corpus *German Web 2020*

Le dieci parole sono ordinate in una lista (*Rangliste*) estratta in base al punteggio (*score*), che ne classifica la tipicità. Dalla lista possono essere ricavate le seguenti informazioni:

- il numero di occorrenze nel corpus di analisi (*focus*);
- il numero di occorrenze nel corpus usato come comparazione (*reference*);
- il punteggio (*score*). *Sketch Engine* calcola il *keyness score* per mezzo di *simple maths*<sup>23</sup> (Kilgarriff 2009).

In particolare, se si prende come esempio il lemma in prima posizione, *Trump*, si vede che ha lo *score* più alto (20.4), che occorre 533.400 volte nel corpus di analisi (*focus*) e 5.862 volte nel corpus di riferimento (*reference*). Andando a cliccare sui tre puntini alla fine della riga possono essere visualizzate le concordanze della parola *Trump* sia nel corpus di analisi, sia in quello di riferimento.

La lista delle *keywords* singole permette di soffermarsi sulle parole tipiche e uniche di un corpus rispetto a un altro, per formulare varie ipotesi. Dalla Figura 2.6 si può intuire, per esempio, che *Donald Trump* e la pandemia da Covid-19 rappresentano due tematiche particolarmente rilevanti nel corpus di analisi.

<sup>23</sup> *Simple Math* è un calcolo statistico utilizzato da *Sketch Engine* per identificare le *keywords* e compara le frequenze del corpus di analisi (*focus corpus*) con le frequenze nel corpus di riferimento (*reference corpus*). Si veda <https://www.sketchengine.eu/documentation/simple-maths/#:~:text=Simple%20maths%20is%20the%20keyness,different%20corpora%20can%20be%20used> [17.1.2023].

La lista delle *keywords* multiple (Figura 2.7), invece, estrae le sequenze di parole (*multi-word expressions*) del corpus di analisi rispetto al corpus di riferimento (*German Web 2013*).

	Term	Frequency <sup>?</sup>		Score <sup>?</sup>	
		Focus	Reference		
1	letztes Jahr	1,457,224	< 18	70.4	...
2	erstes Mal	1,038,947	< 18	50.5	...
3	vergangenes Jahr	961,864	< 18	46.8	...
4	Million Euro	898,800	< 18	43.8	...
5	lange Zeit	866,957	< 18	42.3	...
6	kurze Zeit	827,451	< 18	40.4	...
7	New York	804,843	< 18	39.3	...
8	weite Information	781,822	< 18	38.2	...
9	nahes Jahr	629,036	< 18	31.0	...
10	andere Seite	612,986	< 18	30.2	...

Figura 2.7: *Keywords* multiple del corpus *German Web 2020*

Nella lista delle *keywords* multiple (Figura 2.7) si possono trovare combinazioni tipiche di vario tipo, classificabili in base all'interesse (livello sintattico, semantico, pragmatico ecc.) e all'orientamento teorico scelto. Il termine di paragone è sempre il corpus di riferimento. Dalla Figura 2.7 si può dedurre che il tema ruota soprattutto intorno a eventi passati (*letztes Jahr*, *vergangenes Jahr*) e accaduti per la prima volta (*erstes Mal*).

Un'altra lista ricavabile da un corpus è quella dei *cluster* o segmenti ripetuti (*n-grams*)<sup>24</sup>, ossia di quelle sequenze di 1,2,3...*n* fonemi, grafemi, token che si ripetono identiche all'interno di un corpus (cfr. Spina 2001: 143). Grazie all'analisi dei *cluster* si possono analizzare fenomeni visibili a livello grammaticale, semantico e pragmatico. Si possono focalizzare quindi fraseologismi (cfr. Burger 2015), così come sequenze stereotipate con funzione pragmatica appartenenti ad ambiti tematici particolari, come per esempio l'uso di *ist schlichtweg falsch* usato spesso dal partito tedesco della CDU (*Christlich Demokratische Union*) con la funzione di *Verteidigung* (difesa) (cfr. Bubenhofer 2017: 81). In *Sketch Engine* per l'estrazione dei *cluster* si può utilizzare la funzionalità *N-Gram*.

Se si estrae la lista degli *n-grams* del corpus *German Web 2020* (Figura 2.8), si possono identificare:

24 A seconda dell'approccio teorico e della definizione di collocazione, gli *n-grams* possono essere definiti collocazioni della lunghezza *n* (cfr. Bubenhofer 2017: 70).

- semplici sequenze frequenti di token (*es sich um, in den letzten*), alcune delle quali possono essere completate aumentando il numero di parole (*in den letzten* → *in den letzten Jahren, in den letzten Tagen*);
- combinazioni con vario grado di idiomaticità (*in diesem Jahr, wie z. B., in der Regel, auf jeden Fall*).

N-gram	Frequency <sup>?</sup>
1 in der Regel	71,552 ***
2 in den letzten	67,640 ***
3 es sich um	64,157 ***
4 in diesem Jahr	57,488 ***
5 handelt es sich	54,666 ***
6 und in der	51,890 ***
7 wie z. B.	51,451 ***
8 sich in der	49,299 ***
9 auf jeden Fall	48,353 ***
10 im Rahmen der	48,342 ***

Figura 2.8: N-grams (3-4) del corpus *German Web 2020*

## 2.6 Profilo di parola

Molti sistemi di gestione e analisi permettono l'estrazione di profili di parola, con i quali si sintetizzano le relazioni più frequenti di una particolare unità lessicale: nel DWDS sono chiamati *Wortprofile*, in *Sketch Engine* sono chiamati *Word Sketch*.

Di seguito, il focus è su *Word Sketch* che permette di osservare il comportamento grammaticale e combinatorio dell'unità oggetto d'interesse in un corpus specifico. Questa funzionalità è utile soprattutto in corpora molto estesi, dal momento che ricavare informazioni sul comportamento di un'unità lessicale attraverso le linee di concordanza è praticamente impossibile se il numero delle attestazioni è elevato. Con *Word Sketch* i risultati sono organizzati per relazioni grammaticali, le cui regole sono illustrate nella *sketch grammar*<sup>25</sup> del sistema. Per ogni relazione grammaticale, le combinazioni più tipiche sono quelle in cima alla lista.

Se si prende come esempio il profilo di parola di *Klima* (Figura 2.9) si può osservare, per esempio, con quali verbi *Klima* è utilizzato come soggetto (*verbs with "Klima" as subject*), con quali verbi è utilizzato come complemento oggetto

25 Cfr. [https://www.sketchengine.eu/my\\_keywords/word-sketch-grammar](https://www.sketchengine.eu/my_keywords/word-sketch-grammar) [28.08.23].

(*verbs with "Klima" as object*), con quali aggettivi si combina (*modifiers with "Klima"*) eccetera.

verbs with "Klima" as subject			verbs with "Klima" as accusative object			modifiers of "Klima"		
<b>herrschen</b>	4,510	8.7 ...	<b>vergiften</b>	1,456	8.9 ...	<b>mild</b>	17,038	10.5 ...
Klima herrscht			Klima vergiften			milde Klima		
<b>begünstigen</b>	441	7.8 ...	<b>herrschen</b>	3,971	8.7 ...	<b>mediterran</b>	7,112	9.6 ...
Klima begünstigt			Klima herrscht			mediterranes Klima		
<b>wandeln</b>	685	7.8 ...	<b>schonen</b>	1,843	8.3 ...	<b>tropisch</b>	6,209	9.2 ...
Klima wandelt sich			Klima schonen			tropisches Klima		
<b>erwärmen</b>	352	7.8 ...	<b>schützen</b>	5,614	7.9 ...	<b>gemäßigt</b>	4,229	9.0 ...
Klima erwärmt			Klima schützen			gemäßigtes Klima		
<b>kühlen</b>	272	7.3 ...	<b>retten</b>	3,166	7.6 ...	<b>subtropisch</b>	3,555	8.9 ...
Klima kühlt			das Klima retten			subtropisches Klima		
<b>verändern</b>	1,113	7.2 ...	<b>belasten</b>	1,418	7.5 ...	<b>angenehm</b>	8,790	8.5 ...
Klima verändert sich			belastet das Klima			ein angenehmes Klima		
<b>ändern</b>	1,520	7.1 ...	<b>beeinflussen</b>	2,852	7.2 ...	<b>kontinental</b>	2,448	8.4 ...
Klima ändert sich			Klima beeinflussen			kontinentales Klima		
<b>verschlechtern</b>	207	6.6 ...	<b>schädigen</b>	582	6.8 ...	<b>rau</b>	2,574	8.2 ...
verschlechtert sich das Klima			Klima schädigen			raue Klima		
<b>schonen</b>	153	6.6 ...	<b>gewöhnen</b>	649	6.7 ...	<b>feucht</b>	3,411	8.0 ...
schont das Klima			Klima gewöhnt			feuchtes Klima		
<b>vergiften</b>	104	6.4 ...	<b>vertragen</b>	506	6.6 ...	<b>ganzjährig</b>	1,838	7.8 ...
vergiftet das Klima			Klima vertragen			ganzjährig milde Klima		
<b>vorherrschen</b>	102	6.2 ...	<b>regulieren</b>	369	6.4 ...	<b>warm</b>	6,260	7.7 ...
Klima vorherrscht			regulieren das Klima			warmes Klima		
<b>kippen</b>	150	6.1 ...	<b>heizen</b>	302	6.4 ...	<b>prima</b>	1,434	7.5 ...
das Klima kippt			heizt das Klima			Prima Klima		

Figura 2.9: *Word Sketch* della parola *Klima* nel corpus *German Web 2020*

Prima di osservare nel dettaglio le combinazioni lessicali della Figura 2.9, è opportuna una riflessione: dalla figura si evince, infatti, che la combinazione *Klima* + *herrschen* compare nelle prime posizioni sia della lista *verbs with "Klima" as subject* (4510 occorrenze), sia della lista *verbs with "Klima" as object* (3.971 occorrenze). È necessario quindi visualizzare le concordanze delle singole combinazioni cliccando sui tre puntini alla fine della riga e scegliendo l'opzione<sup>26</sup> *Concordance (Klima + herrschen)*. Nelle figure 2.10 e 2.11 sono visibili le prime dieci concordanze:

26 Oltre a *Concordance*, dai tre puntini si può accedere a *Multivords Sketch* (per esempio *subtropisch + Klima + herrschen*), che permette di allargare il contesto delle combinazioni tipiche estratte in precedenza, e a *Thesaurus* (per esempio, il primo verbo che ha un profilo simile è *scheinen*).

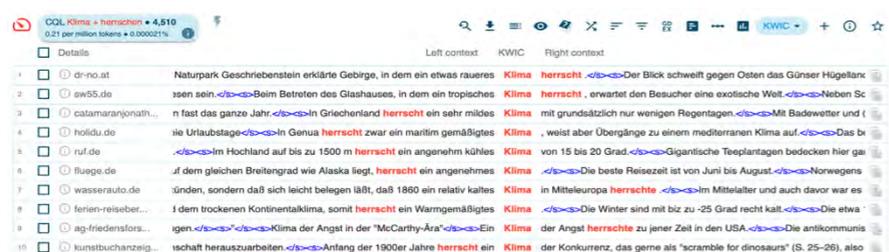


Figura 2.10: Visualizzazione della combinazione *Klima* + *herrschen* dalla colonna *verbs with "Klima" as subject* nel corpus *German Web 2020*

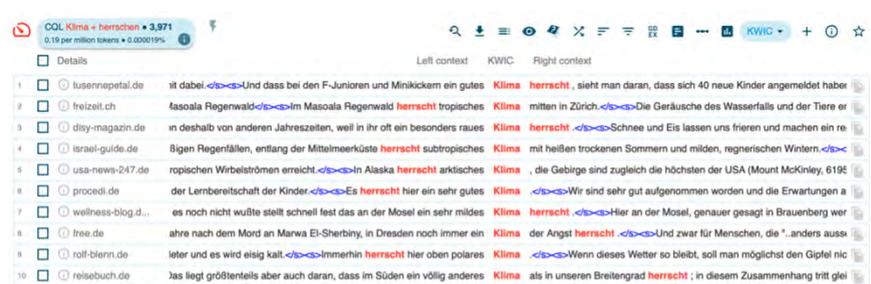


Figura 2.11: Visualizzazione della combinazione *Klima* + *herrschen* dalla colonna *verbs with "Klima" as object* nel corpus *German Web 2020*

Se si osservano la frase n. 3 della Figura 2.10 (1)<sup>27</sup> e la frase n. 2 della Figura 2.11 (2), appare evidente che in entrambe *Klima* è il soggetto del verbo *herrschen*:

- (1) In Griechenland **herrscht**<sup>28</sup> ein sehr mildes **Klima** mit grundsätzlich nur wenigen Regentagen.
- (2) Im Masoala Regenwald **herrscht** tropisches **Klima** mitten in Zürich.

In realtà, in tutte le frasi delle due categorie *verbs with "Klima" as object* e *verbs with "Klima" as subject*, la parola *Klima* ha sempre la funzione di soggetto. Si può dunque concludere che l'estrazione automatica di combinazioni lessicali (profili di parola ecc.), indispensabile per molti tipi di analisi, deve essere sempre supportata da una attenta analisi qualitativa.

Una volta escluse le relazioni classificate erroneamente, si passa ad analizzare il profilo di parola di *Klima* (Figura 2.9)<sup>29</sup>.

27 Salvo diversa indicazione, tutti gli esempi sono estratti da corpora e sono fedeli all'originale. Eventuali errori e/o omissioni non sono stati corretti.

28 Il grassetto, salva diversa indicazione, è ad opera dell'autrice.

29 Per motivi di sintesi sono state mostrate solo alcune delle relazioni identificate tramite *Word Sketch*.

Per ogni combinazione elencata possono essere estratte informazioni riguardo al numero di occorrenze della combinazione nel corpus e al suo punteggio (*score*), calcolato con il coefficiente *logDice*<sup>30</sup>. Se si prende come esempio la prima attestazione della colonna *verbs with "Klima" as subject* (*Klima + herrschen*), si vede che ha lo *score* più alto (8.7) e un numero di occorrenze pari a 4.510. Le combinazioni sono, infatti, ordinate per numero decrescente di *score*: in cima alla lista si trovano le più tipiche, che non sono sempre però le più frequenti (è possibile, però, estrarre le relazioni usando anche il criterio della frequenza). Quindi, le combinazioni tipiche della relazione *verbs with "Klima" as subject* sono *Klima + herrschen*, *Klima + begünstigen*, *Klima + wandeln*, ecc.; quelle della relazione *verbs with "Klima" as object* sono *Klima + vergiften*, *Klima + schonen*, *Klima + schützen*; quelle della relazione *modifiers with "Klima"* sono *mild + Klima*, *mediterran + Klima*, *tropisch + Klima*.

Il profilo di parola, come altre funzionalità, permette di visualizzare le combinazioni più tipiche in un corpus. Se finora si è optato per l'uso del termine neutro "combinazione lessicale", risulta ora necessario introdurre il termine "collocazione". Le collocazioni sono un tipo specifico di combinazione<sup>31</sup>, ma con varie possibilità di definizione e interpretazione in base all'orientamento teorico, e raggruppabili in due filoni: uno empirico (1) e uno teorico (2) (cfr. anche Flinz 2019: 92).

1. Il filone empirico viene generalmente associato a John Sinclair e a John Rupert Firth<sup>32</sup>. Nel suo lavoro del 1991, John Sinclair mostra come il linguaggio sia caratterizzato da due principi fondamentali, che determinano la produzione di enunciati dotati di significato: il principio di scelta aperta (*open-choice principle*), per il quale ogni enunciato viene prodotto tramite un complesso sistema di scelte a vari livelli (suono, parole, sintagmi ecc.) guidate dal criterio di grammaticalità, e il principio idiomatico (*idiom principle*), secondo il quale gli elementi linguistici tendono a co-occorrere e a costituire una scelta singola, un blocco unico. Il principio idiomatico determina la tendenza delle parole ad attrarre nella loro immediata vicinanza altre forme specifiche, ossia le loro collocazioni (cfr. Spina 2001: 136). In evidenza vi è quindi l'importanza delle ricorrenze lessico-grammaticali per l'identificazione di pattern strutturali e semantici. Grazie a John Rupert Firth e all'uso dell'informatica e dei metodi statistici applicati a corpora di grandi

30 Cfr. [https://www.sketchengine.eu/my\\_keywords/logdice/](https://www.sketchengine.eu/my_keywords/logdice/) [17.1.2023]. Il coefficiente *logDice* permette l'identificazione delle combinazioni più tipiche e quindi più "forti". Il coefficiente *logDice* non è influenzato dalla dimensione del corpus. Per una panoramica di altri parametri statistici, cfr. Stefanowitsch (2020: 224-233).

31 Le collocazioni sono unità stereotipate, che si distinguono però da altri fraseologismi come proverbi e modi dire (cfr. Burger 2015: 38).

32 Nel mondo germanofono si citano anche i lavori di Evert (2009), che distingue tra collocazioni e *multivords units* (*Mehrwortverbindungen*), e Steyer (2013), che introduce il termine "combinazione usuale" (*usuelle Wortverbindung*).

dimensioni ha preso, invece, avvio lo studio empirico delle collocazioni. Le collocazioni vengono estratte proprio per mezzo di parametri statistici che determinano quali combinazioni sono più frequenti in un determinato segmento di testo di dimensione fissa (*Textfenster* in tedesco, *span* in inglese), indipendentemente da fattori di tipo casuale (cfr. Lemnitzer/Zinsmeister 2015: 179). I metodi statistici<sup>33</sup>, mettendo in relazione la frequenza dei collocati all'interno dello *span* e dell'intero corpus, misurano proprio la significatività di una collocazione. Tali collocazioni possono poi essere organizzate per relazioni di vario tipo, semantico, sintattico eccetera.

2. Per il filone teorico si distinguono da un lato la definizione di Hausmann (1984: 401f), per il quale le collocazioni sono unità lessicali in relazione sintattica (di solito sono combinazioni di due parole che si trovano in un *Determinationsverhältnis* come ad esempio *schütteres Haar*, dove *Haar* è la base e *schütter* è il collocatore)<sup>34</sup>, e, dall'altro, quella di Burger (2015: 38-41), per il quale le collocazioni sono un tipo di fraseologismo. Burger distingue varie categorie:

- fraseologismi verbali con oggetto/oggetti (*Verbale Phraseme mit Objekt(en)*) come *Sport treiben*;
- fraseologismi verbali con avverbiale + oggetto/oggetti (*Verbale Phraseme mit Adverbiale (+Objekt(en))*) come *etw. Online bestellen*;
- fraseologismi verbali con complementi complessi (*Verbale Phraseme mit komplexeren Ergänzungen*), come *bei rot über die Ampel fahren*;
- fraseologismi nominali (*Nominale Phraseme*), come *persönliche Daten*;
- fraseologismi aggettivali (*Adjektivische Phraseme*), come *bitter nötig*.

L'analisi delle collocazioni è un presupposto fondamentale per comprendere fenomeni legati sia alla sfera lessico-grammaticale, sia alla sfera semantica (cfr. Spina 2001: 134). Indipendentemente dall'orientamento teorico, l'approccio empirico resta, in ogni caso, un buon punto di partenza.

Le applicazioni dello studio delle collocazioni possono essere varie, spaziando dalla lessicografia alle scienze della traduzione fino alla didattica delle lingue straniere.

33 Per l'estrazione di collocazioni si possono utilizzare vari parametri statistici. Oltre al già citato *logDice*, usato da *Sketch Engine*, vi sono anche il *Mutual Information* (cfr. Spina 2001: 138-140; Stefanowitsch 2020: 226-227), il *T-Score* (cfr. Spina 2001: 138-143); il *Chi-Square* (cfr. Stefanowitsch 2020: 225-226), il *log-likelihood ratio test* (cfr. Stefanowitsch 2020: 225-227), il *Minimum Sensitivity* (cfr. Stefanowitsch 2020: 228), il *Fisher's exact test* (cfr. Stefanowitsch 2020: 228-229). Per un confronto dei parametri in base alle caratteristiche del corpus e alle possibili finalità cfr. Stefanowitsch (2020: 229-233). Spina (2001: 140) si sofferma su *Mutual Information* e *T-Score*.

34 Si veda anche la definizione di Feilke (1996), che amplia la definizione di Hausmann e pone l'attenzione sulla relazione lessico-grammatica.

## 2.7 Altre possibilità di analisi

In *Sketch Engine* sono presenti due ulteriori funzionalità interessanti per vari tipi di applicazione. Una di queste è *Word Sketch Difference*, che confronta il profilo di parola (*word sketch*) di due parole. Si prendano come esempio le due parole *Klima* e *Umwelt*. Come si evince dalla Figura 2.12, il confronto avviene relazione per relazione: *verbs with "Klima" as subject*, *verbs with "Klima" as object*, *modifiers with "Klima"* eccetera.

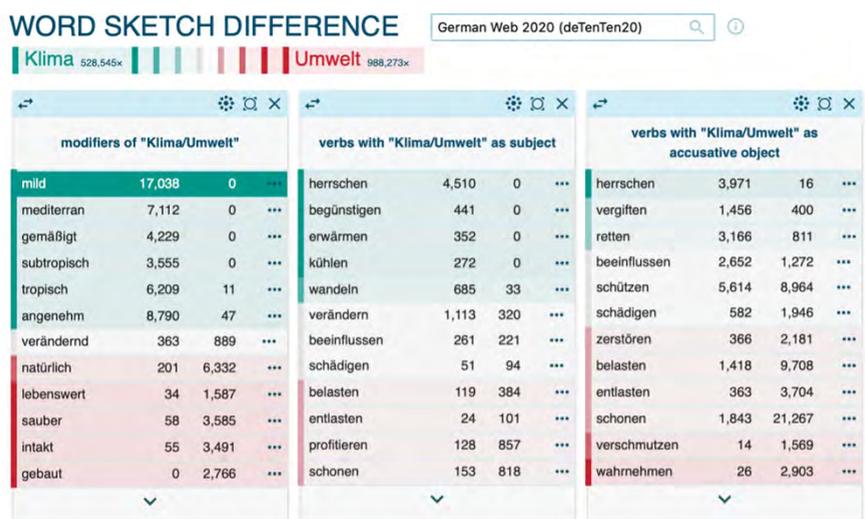


Figura 2.12: *Word Sketch Difference* delle parole *Klima* e *Umwelt* nel corpus *German Web 2020*

In verde nitido (come la parola *Klima*) sono evidenziati tutti gli aggettivi maggiormente ricorrenti con *Klima* (*mild*, *mediterran* ecc.), mentre in rosso nitido (come la parola *Umwelt*) tutti gli aggettivi più usati con *Umwelt* (*lebenswert*, *sauber* ecc.). Più il colore perde di nitidezza, meno la combinazione risulta tipica. In grigio sono colorate, invece, le parole che possono essere usate con entrambe (*verändernd*). Questa funzionalità si presta per confronti non solo a livello sintattico, ma anche a livello semantico.

Un'altra funzionalità presente in *Sketch Engine* è *Thesaurus*. Con *Thesaurus* si può generare automaticamente una lista di sinonimi (o di parole che si avvicinano a questa relazione semantica) rispetto a una parola data. Le parole estratte apparterranno quindi al medesimo campo semantico; secondo la semantica distribuzionale<sup>35</sup>, infatti, le parole che possono essere usate nel medesimo contesto sono simili anche dal punto di vista del significato.

35 Cfr. <https://www.sketchengine.eu/guide/thesaurus-synonyms-antonyms-similar-words/#toggle-id-3> [24.08.23].

*Sketch Engine* utilizza i *word sketches* delle parole per determinare il loro grado di somiglianza, consigliando però l'uso di corpora di una certa grandezza (minimo 100.000 parole) per avere risultati più affidabili. Più grande sarà la dimensione del corpus, maggiore sarà la probabilità che i risultati siano corretti.

Se si estrae la lista delle parole affini dal punto di vista semantico a *Klima* dal corpus *German Web 2020* (Figura 2.13), nelle prime posizioni si possono identificare le parole più simili alla parola *Klima*, ossia *Umwelt*, *Kultur*, *Atmosphäre*. Segue il dato della frequenza nel corpus e il punteggio che ne determina la similarità (numero che indica la percentuale delle combinazioni in comune):

	Lempos	Frequency ?	Similarity ? ↑
1	Umwelt	988,273	0.330 ...
2	Kultur	1,867,107	0.307 ...
3	Atmosphäre	802,264	0.300 ...
4	Natur	2,142,253	0.297 ...
5	Klimawandel	330,599	0.293 ...
6	Wetter	1,221,088	0.291 ...
7	Gesundheit	1,290,367	0.290 ...
8	Vielfalt	716,701	0.289 ...
9	Verhältnis	1,530,086	0.282 ...
10	Landschaft	846,173	0.279 ...

**Figura 2.13:** *Thesaurus* della parola *Klima* nel corpus *German Web 2020*

Cliccando sui tre puntini accanto a ogni parola si potrà scegliere tra quattro funzionalità; se si prende come esempio *Umwelt*, si potrà:

- utilizzare *Word Sketch Difference* per confrontare *Klima* con *Umwelt*;
- consultare il *Thesaurus* della parola *Umwelt*;
- identificare le concordanze della parola *Umwelt* tramite *Concordance*;
- estrarre il profilo di parola di *Umwelt* tramite *Word Sketch*.

# Capitolo 3

## Corpora in lingua tedesca

I corpora con dati primari attualmente disponibili in lingua tedesca sono numerosi. Possono essere monolingui o plurilingui, corpora di riferimento o specialistici, corpora grezzi o annotati eccetera. Per poter capire, però, quale corpus possa risultare più adatto al proprio interesse di ricerca, una classificazione mediante criteri precisi può risultare utile per orientarsi. Nella prima parte di questo capitolo è presentata la classificazione proposta da Lemnitzer/Zinsmeister (2015: 137), rimodellata però in alcuni punti. Oggetto di discussione sono, inoltre, i criteri adottati per la sua realizzazione, con il supporto e l'integrazione di ulteriore letteratura (3.1). Segue la presentazione dei tre corpora principali per la lingua tedesca, con focus sulle loro caratteristiche peculiari, nonché sugli strumenti che ne consentono la consultazione (3.2). Il capitolo si conclude con un breve excursus su altri tipi di corpora, offrendo una panoramica delle loro caratteristiche principali.

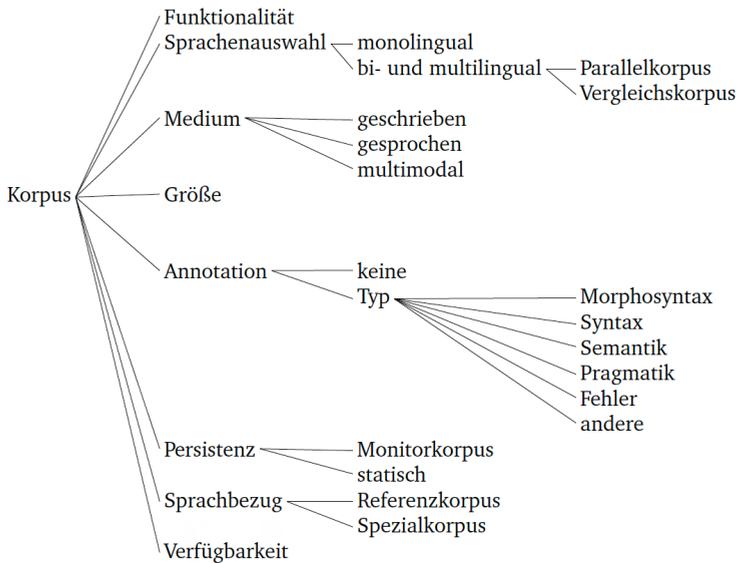
### 3.1 Classificare i corpora

Il panorama dei corpora presenti nel web è assai variegato<sup>1</sup>. Non è facile orientarsi e trovare immediatamente quello più adatto ai propri scopi. Per capire, quindi, quali corpora possano essere adatti a un obiettivo di ricerca prefissato, è necessario essere a conoscenza delle loro caratteristiche principali. Di seguito viene dunque presentata e discussa una classificazione utile a distinguere i vari corpora presenti in rete, quella di Lemnitzer/Zinsmeister (2015: 137), che, pur rappresentando soltanto una delle possibili classificazioni, è stata scelta per via della sua strutturazione chiara e intuitiva (Figura 3.1).

Secondo Lemnitzer/Zinsmeister (2015) i corpora possono essere classificati in base ai seguenti criteri: obiettivo per il quale il corpus è stato costruito (*Funktionalität*), lingua dei dati primari (*Sprachenauswahl*), canale di trasmissione dei dati (*Medium*), grandezza (*Größe*), annotazione (*Annotation*), persistenza (*Persistenz*), tipo di relazione con la lingua oggetto di studio (*Sprachbezug*), disponibilità (*Verfügbarkeit*).

---

1 Una panoramica aggiornata è fornita dal progetto europeo CLARIN (<https://www.clarin.eu>) [21.10.2022], nato con l'obiettivo di realizzare un *Virtual Language Observatory* aggiornato con corpora in varie lingue europee. Cfr. <https://vlo.clarin.eu/?1> [23.10.22].



**Figura 3.1:** Classificazione di corpora (Lemnitzer/Zinsmeister 2015: 137)

Prima di procedere a una riflessione dettagliata dei criteri sopraelencati, è opportuno sottolineare il fatto che essi, in realtà, possono essere anche raggruppati in due sottocategorie: da un lato, i criteri riguardanti il design del corpus (*Designkriterien*), che fanno riferimento alle principali caratteristiche dei dati primari (la lingua dei dati, il canale di trasmissione dei dati, la grandezza del corpus, il tipo di relazione dei dati con la lingua oggetto di studio, l'obiettivo per cui è stato costruito il corpus); dall'altro, quelli concernenti la preparazione del corpus (l'annotazione) (cfr. Lemnitzer/Zinsmeister 2015: 138). Inoltre, i criteri possono essere anche combinati tra di loro: un corpus può, cioè, essere monolingue, ma anche specialistico, annotato eccetera.

Il primo criterio, l'obiettivo per il quale è stato costruito o si costruisce il corpus (*Funktionalität*), è quello fondamentale, perché influisce su tutti gli altri, anche se può essere riadattato in corso d'opera. Normalmente i corpora vengono costruiti per fungere da base empirica per vari tipi di analisi con obiettivi più o meno specifici, come per esempio la documentazione linguistica (il corpus *Emigrantendeutsch in Israel*), per fini lessicografici (il *DWDS-Kernkorpus*), per studi sull'apprendimento e per l'insegnamento di una lingua (i corpora Falko e MERLIN), per studi sulla lingua usata per comunicare in Internet (*internetbasierte Kommunikation*) come il *Mobile Communication Database 2* (MoCoDa 2)<sup>2</sup>.

Il secondo criterio, la lingua (*Sprachenauswahl*), riguarda la lingua/le lingue dei dati primari del corpus. La prima distinzione che deve essere fatta è tra corpora

2 Per una breve descrizione di questi corpora si veda il paragrafo 3.3.

monolingui (*monolinguale Korpora*) in cui i dati sono disponibili in una sola lingua<sup>3</sup> e corpora bilingui o multilingui (*bilinguale oder multilinguale Korpora*) in cui i dati sono accessibili in due o più lingue. I corpora bilingui o multilingui, a loro volta, si distinguono in corpora paralleli<sup>4</sup> (*Parallelkorpora*) e corpora comparabili<sup>5</sup> (*Vergleichskorpora*). Nel primo caso, quello dei corpora paralleli (*Parallelkorpora*), vi sono dati appartenenti a una L1 e la loro traduzione in una o più lingue (L2, L3 ... Ln) (Lemnitzer/Zinsmeister 2015: 138). Essi possono essere sia monodirezionali che bidirezionali (cfr. Freddi 2019: 70-71). Nel secondo caso, quello dei corpora comparabili (*Vergleichskorpora*), vi sono testi in una L1 e altri testi confrontabili per genere, registro, funzione comunicativa in altre lingue (L2, L3 ... Ln), che non sono però traduzioni.

Il terzo criterio, il canale di comunicazione (*Medium*), ossia la variazione dialettale, permette di distinguere tra corpora di lingua scritta, con dati che sin dalla loro origine sono stati prodotti per essere scritti<sup>6</sup> (quindi non dati orali che sono stati trascritti), e corpora di lingua parlata, che raccolgono dati audio e video, da trascrivere ai fini dell'analisi<sup>7</sup>. Sono necessarie però alcune precisazioni: innanzitutto, le trascrizioni, di fatto il primo stadio di annotazione di testi orali, implicano scelte teoriche e operative (cfr. Freddi 2019: 53); le scelte dovrebbero essere tali da non ridurre la complessità del dato orale e da permettere un'interpretazione adeguata dei fenomeni studiati. In secondo luogo, le trascrizioni sono un'interpretazione di un dato audio<sup>8</sup>. I due livelli dovrebbero essere op-

3 Nei corpora monolingui sono presenti sia la lingua standard, sia le sue varianti (diatopiche, diastratiche, diacroniche ecc.).

4 Corpora paralleli sono il corpus OPUS (*Open Parallel Corpus*) (<https://opus.nlpl.eu>) [21.10.2022], che in realtà è un insieme di testi tradotti presi da Internet e il corpus EUROPARL (*European Parliament Proceedings Parallel Corpus*) (<https://www.statmt.org/europarl/>) [21.10.2022].

5 Si consideri che il termine *Vergleichskorpus* ha nelle scienze della traduzione un significato diverso. In tal caso si tratta di un corpus che contiene sia testi originari (L1), sia le sue traduzioni (L1, L2 ecc.).

6 Casi limite sono per esempio i protocolli stenografati dei dibattiti parlamentari (cfr. Lemnitzer 2022: 412).

7 Oggi la trascrizione può essere fatta mediante sistemi di riconoscimento vocale (cfr. f4x, <https://www.audiotranskription.de>). Un controllo manuale, per quanto i sistemi automatici diventino sempre più precisi, è comunque consigliabile.

8 Per la trascrizione cfr. Ochs (1979). Che cosa viene trascritto e come lo si trascrive dipende dall'obiettivo della ricerca. Sono possibili vari tipi di trascrizione (cfr. anche Morek 2022: 391ff): una possibilità è la trascrizione ortografica; un'altra è costituita dalla trascrizione prosodica, in cui si tiene conto di fenomeni tipici del parlato quali l'intonazione, il ritmo, l'accento ecc. Per il tedesco si è imposto il seguente sistema di trascrizione: il *Gesprächsanalytische Transkriptionssystem* (GAT), oggi nella versione aggiornata e rivista GAT2 (cfr. Hagemann/Henle 2021), che permette di realizzare trascrizioni minime (*Minimaltranskripte*) e trascrizioni di base (*Basistranskripte*). La trascrizione può essere fatta con diversi sistemi, tra cui FOLKER, un editor che comprende anche un player audio e permette varie visualizzazioni della trascrizione (cfr. <https://agd.ids-mannheim.de/folker.shtml>) [21.10.2022]. FOLKER utilizza un sistema cGAT (c = *computergestützt*), una versione leggermente modificata di GAT2.

portunatamente separati, anche se oggi, con l'allineamento tra testo e audio, gli studi possono raggiungere un maggior grado di completezza (cfr. Freddi 2019: 53). I corpora di lingua orale possono presentare, inoltre, diversi gradi di oralità e spontaneità, dovuti al fatto che i soggetti parlanti possono essere persone che dialogano spontaneamente, professionisti opportunamente preparati, oppure persone che semplicemente leggono un testo scritto. Il corpus può contenere quindi dati con un diverso grado di oralità (cfr. Koch/Oesterreicher 1994). Nelle pubblicazioni più recenti si parla di un ulteriore tipo di corpus, ossia il corpus della comunicazione in Internet (cfr. Beißwenger/Lüngen 2022: 433). Anche se le sue caratteristiche sono a cavallo tra lingua scritta e lingua parlata, per la sua costruzione e il suo utilizzo sono necessarie convenzioni, procedure e strumenti *ad hoc* (cfr. Storrer 2014: 189-191). I corpora multimodali contengono, invece, anche dati video e sono indispensabili per l'analisi della gestualità o anche del linguaggio dei segni.

Il criterio della grandezza (*Größe*) si riferisce al numero di parole o token del corpus. Come già anticipato, la prima generazione di corpora aveva un numero limitato di token (ca. un milione) rispetto ai corpora moderni, che invece superano anche il miliardo di token, come il DWDS-*Zeitungskorpus* (23 miliardi), DeReKo (66,6 miliardi) e i *web corpora* di *Sketch Engine* (i cosiddetti corpora *TenTen*). La grandezza rappresenta un criterio fondamentale quando si vogliono fare analisi di tipo quantitativo per mezzo di parametri statistici. Per estrarre, ad esempio, i profili di parola (*Wortprofile*), il corpus deve essere abbastanza grande da poter estrarre un numero considerevole di lessemi co-occorrenti. Solo così possono essere determinati i potenziali candidati che andranno a completare il profilo della parola ricercata. La grandezza di un corpus è un criterio rilevante anche per la ricerca di fenomeni rari o complessi: si pensi ad analisi *corpus-based* (in tedesco *korpusgestützt*, si veda il paragrafo 1.3), che hanno come obiettivo la verifica di una determinata ipotesi. La grandezza di un corpus non deve essere però l'unico criterio da utilizzare per la scelta di un corpus, che deve essere dettata, invece, dall'obiettivo della ricerca. Anche con corpora piccoli si possono fare, infatti, ricerche valide, soprattutto per analizzare le lingue di specialità (cfr. Flinz 2018; Flinz/Perkuhn 2018) o un determinato discorso<sup>9</sup> (cfr. Brambilla/Flinz 2020; Flinz 2021b; Flinz 2022).

Il criterio dell'annotazione (*Annotation*) permette di distinguere in prima istanza tra corpora grezzi, non annotati, e corpora annotati. I corpora annotati si differenziano, inoltre, principalmente in corpora annotati a livello morfosintattico, sintattico, semantico e pragmatico (cfr. paragrafo 1.2). Vi sono però anche altri tipi di annotazione, che riguardano aspetti quali la struttura testuale, la struttura informativa, la prosodia eccetera.

9 Per discorso si intende un corpus virtuale realizzato in base a criteri semantici (cfr. Busse/Teubert 1994: 14).

Il criterio della persistenza<sup>10</sup> (*Persistenz*) consente di differenziare, da un lato, i corpora statici (*statische Korpora*), ossia quei corpora che sono stati creati in un arco temporale definito e che vengono salvati in un determinato stadio (i corpora statici possono essere poi anche ampliati; in tal caso si avranno semplicemente delle nuove versioni del corpus); dall'altro, i corpora di monitoraggio (*Monitorkorpora*), che vengono aggiornati periodicamente (su base annuale, mensile e, in alcuni casi, quotidiana), senza l'alterazione degli altri parametri, come il bilanciamento dei testi (un esempio è il corpus *Webmonitor*)<sup>11</sup>. Questi tipi di corpora si prestano molto bene per lo studio del cambiamento della lingua e per l'identificazione di neologismi. Un altro tipo di corpus sono i corpora dinamici. Occorre però precisare che non tutte/i le/gli studiose/i sono concordi nell'introdurre questa terza categoria. Mentre per Lemnitzer/Zinsmeister (2015) e Freddi (2019) i corpora di monitoraggio e i corpora dinamici sono la stessa cosa, Perkuhn/Keibel/Kupietz (2012: 51) distinguono i corpora di monitoraggio da quelli dinamici: quelli di monitoraggio (*Monitorkorpora*) hanno sempre la medesima grandezza anche se i dati cambiano, quelli dinamici (*dynamische Korpora*) sono invece corpora che vengono incrementati e modificati. Un esempio di corpus dinamico è il corpus di riferimento DeReKo, che viene ampliato continuamente e a intervalli regolari (per citarlo deve essere fatto riferimento alla versione, per esempio alla data del 31.03.23 è fruibile DeReKo-2023-I). Collegato alla persistenza è anche il tema della riproducibilità (*Reproduzierbarkeit*) delle analisi in un corpus. Con i corpora statici non ci sono problemi, ma in caso di corpora di monitoraggio o di corpora dinamici i dati cambiano; i risultati che si ottengono con la stessa procedura di analisi potrebbero quindi variare.

Il criterio del tipo di relazione con la lingua oggetto di studio (*Sprachbezug*), chiamato criterio della variazione diafasica in Freddi (2019: 44), permette di differenziare tra corpora generici di riferimento di una lingua (*Referenzkorpora*)<sup>12</sup>, che cercano di rappresentare in maniera più fedele possibile una lingua in un determinato periodo di tempo (per esempio il tedesco del secolo X) con tutte le sue varietà diafasiche, diastratiche, diatopiche, diamesiche, e i corpora specialistici (*Spezialkorpora*), che focalizzano un singolo genere o dominio. Nel primo caso, le caratteristiche principali sono l'ampiezza e la varietà delle tipologie testuali, che dovrebbe essere bilanciata (*ausgewogen*)<sup>13</sup>, mentre nel secondo la specificità

10 Il termine "persistenza" ha origine in ambito informatico e indica la possibilità di far sopravvivere i dati all'esecuzione del programma stesso. Cfr. Perkuhn/Keibel/Kupietz (2012: 50-51).

11 Cfr. <https://www.dwds.de/d/korpora/webmonitor> [23.10.22].

12 Il termine *Referenzkorpora* è utilizzato con un altro significato in sistemi di analisi come *Sketch Engine*, dove indica semplicemente il corpus di riferimento con il quale è paragonato il corpus di analisi per identificare le *keywords*.

13 Se si prende come esempio il DWDS-*Kernkorpora* (1900–1999), si vince che la distribuzione dei generi testuali è abbastanza bilanciata, in quanto manualistica e saggistica sono il 21,7 %, i testi letterari il 26,35%, i testi scientifici il 24,59% e quelli giornalistici il 27,29%.

potrebbe riguardare vari aspetti, come la scelta della tipologia testuale o la scelta di un argomento specifico (fenomeni discorsivi ecc.). I corpora di riferimento possono essere usati anche come corpora di controllo (*Kontrollkorpora*), per esempio per analizzare i corpora specialistici. La comparazione dei due corpora permette di mettere in evidenza le somiglianze e le differenze nella distribuzione del fenomeno analizzato (cfr. Flinz/Perkuhn 2018).

Il criterio della disponibilità (*Verfügbarkeit*) riguarda il tipo di accesso al corpus. Un corpus può essere liberamente fruibile tramite una piattaforma o semplicemente scaricabile, oppure può essere consultato previa registrazione o sottoscrizione di una licenza d'uso. Il tipo di disponibilità dipende da molti fattori, quali i diritti di autore, le licenze, i diritti d'uso ecc. (cfr. Perkuhn/Keibel/Kupietz 2012: 52-54).

Nella sua classificazione Freddi (2019: 58-62) inserisce anche altri criteri e discute altre caratteristiche, tra cui il criterio della variazione diacronica per distinguere i corpora sincronici, che offrono uno spaccato di una lingua o di una sua varietà in un preciso arco temporale, e i corpora diacronici, in cui i testi sono stati campionati in periodi diversi<sup>14</sup>. Inoltre, distingue tra i corpora e il *World Wide Web*, mostrando come di fatto il Web, nonostante sia la più grande banca dati esistente e anche facilmente accessibile, non sia adatto a ricerche linguistiche di tipo complesso per una serie di problematiche (dati scelti e organizzati senza nessun criterio specifico, origine non verificabile dei dati, tipologie di testi non sempre facilmente classificabili, volatilità e qualità dei dati ecc.). Motori di ricerca come *Google*, grazie ai loro strumenti di indicizzazione efficienti, possono essere usati per alcuni ambiti della ricerca linguistica, ma allo stato attuale non per ricerche complesse. In tempi recenti sono stati sviluppati sistemi per la realizzazione di *web corpora*, come il *WebBootCat* (cfr. Baroni/Bernardini 2004), e in alcune piattaforme sono stati inseriti *web corpora* già pronti per essere utilizzati, come i corpora *TenTen* in *Sketch Engine*, i *web corpora* nel *Wortschatz Leibniz*<sup>15</sup>, il *Webcorpus* nel *DWDS*<sup>16</sup>.

### 3.2 Principali raccolte di corpora esistenti per la lingua tedesca

In questo paragrafo sono presentate le principali raccolte di corpora per la lingua tedesca e i loro sistemi di gestione e analisi (*Korpusrecherchesysteme*).

14 Si ricorda che lo studio di una lingua in diacronia può essere affrontato anche analizzando corpora sincronici campionati in periodi diversi secondo lo stesso criterio di selezione dei testi.

15 In questa piattaforma non è possibile analizzare i dati in autonomia, ma sono consultabili informazioni già "preconfezionate". Per ogni unità lessicale cercata vengono fornite informazioni come articolo, ambito di appartenenza, descrizione, esempi, parole co-occorrenti. Cfr. [https://corpora.uni-leipzig.de/de?corpusId=deu\\_news\\_2021](https://corpora.uni-leipzig.de/de?corpusId=deu_news_2021) [21.10.2022].

16 Cfr. <https://www.dwds.de/d/korpora/web> [21.10.2022].

Partendo da alcuni criteri presenti nella tipologia presentata nel paragrafo 3.1, per ogni corpus sono messe in luce le sue caratteristiche principali, gli strumenti di ricerca, le modalità di ricerca e di visualizzazione dei risultati.

### 3.2.1 *Das Deutsche Referenzkorpus (DeReKo)*

DeReKo è la più grande raccolta di corpora per la lingua tedesca del periodo contemporaneo e anche del passato più recente (cfr. Lungen 2017). Appositamente realizzato dal *Leibniz-Institut für Deutsche Sprache* per ricerche di tipo linguistico, è composto da ca. 53 miliardi di parole (DeReKo-2023-I). È sostanzialmente un insieme di archivi di corpora e di corpora singoli monolingui in lingua tedesca. In alcuni corpora, per esempio nei corpora di *Wikipedia* (WP), sono presenti anche dati in altre lingue<sup>17</sup> (inglese, croato, francese, italiano, norvegese, polacco, spagnolo e ungherese).

DeReKo comprende solo testi scritti, anche se in alcuni corpora, come nel *Dortmunder Chat Korpus*, i dati si possono ascrivere al cosiddetto “parlato concettuale”, cioè il medium è scritto, ma la concezione è orale (sono *konzeptionell mündlich*)<sup>18</sup>. I testi, appartenenti a vari generi testuali (testi letterari, scientifici, divulgativi, giornalistici), sono stati scelti in base a criteri come quantità, variazione, qualità e attualità.

DeReKo è stato creato come corpus di riferimento, ma comprende anche corpora storici (come il *Goethe Korpus* e il *Grimm Korpus*), corpora letterari (biografie di autori), corpora giornalistici (*Bonner Zeitungskorpus* ecc.), corpora specialistici (*Fußball-Liveticker*), corpora CMC (*computer mediated communication* come il *Dortmunder Chatkorpus 2.2*) e corpora comparabili (i *Wikipedia Korpora*). Si tratta di un corpus dinamico aggiornato a intervalli di tempo regolari<sup>19</sup> e fruibile gratuitamente previa registrazione; alcuni suoi corpora possono essere anche scaricati.

I dati, codificati nel sistema I5 (un adattamento di TEI), sono annotati con varie modalità e possono essere consultati utilizzando due sistemi di ricerca e analisi: COSMAS II (*Corpus Search Management and Analysis System*) e KorAP (*Korpusanalyseplattform*)<sup>20</sup>.

La ricerca in DeReKo tramite COSMAS II avviene per mezzo di vari passaggi, illustrati qui di seguito:

17 Per la loro natura intrinseca (testi appartenenti al medesimo genere testuale), i corpora di *Wikipedia* possono essere usati anche come corpora comparabili (*Vergleichskorpora*). Cfr. Gredel/Flinz (2020); Flinz/Gredel (2022); Storrer/Herzberg (2022).

18 Cfr. Koch/Oesterreicher (1994).

19 La sua crescita viene stimata intorno ai 3,1 milioni di parole l'anno (cfr. Kupietz/Lungen/Kamocki/Witt 2018: 4353).

20 KorAP è stato realizzato per colmare alcune criticità di COSMAS II, come l'impossibilità di combinare vari livelli di annotazione. KorAP non ha un linguaggio di ricerca proprio, ma si basa su COSMAS II. KorAP è stato integrato anche in altri progetti (cfr. il corpus nazionale della lingua rumena). Per maggiori informazioni, cfr. Lemnitzer/Diewald (2022: 380-381).

(1) per prima cosa occorre scegliere l'archivio o il corpus nel quale si vuole effettuare la ricerca<sup>21</sup>. Nella versione attuale (DeReKo-2023-I) sono presenti 568 corpora organizzati in 18 archivi<sup>22</sup>;

(2) successivamente deve essere scelto il corpus da analizzare. Si può scegliere tra i corpora predefiniti presenti all'interno dell'archivio scelto (per esempio l'archivio W), tra i corpora virtuali dell'utente (se sono stati creati in precedenza) e tra i corpora virtuali messi a disposizione dalla comunità scientifica. Se si opta per la prima opzione (archivio W), si potrà scegliere tra corpora di vario tipo<sup>23</sup>. La possibilità di creare corpora virtuali<sup>24</sup> che rimangono in memoria e possono essere rianalizzati di volta in volta, è una funzionalità e una possibilità molto interessante di COSMAS II, perché permette all'utente di lavorare anche con corpora più piccoli, adatti ai propri interessi di ricerca (cfr. anche Flinz 2021b), e di svolgere analisi qualitative molto accurate;

(3) Una volta scelto il corpus (il corpus *W-öffentlich* dell'archivio W), si può procedere a formulare la ricerca (Figura 3.2).

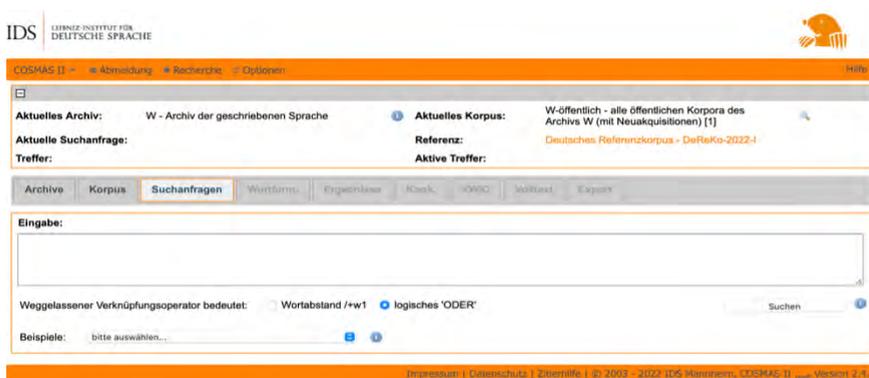


Figura 3.2: Screenshot della maschera di ricerca in COSMAS II

L'utente può scegliere se cercare semplicemente la forma di una parola (una parola coniugata) o la sua forma base (il lemma), oppure una combinazione di più parole o una frase intera. Gli esempi forniti dal sistema (*Beispiele*) permettono

21 Per esempio, W (*Archiv der geschriebenen Sprache*), PARONYM (*Archiv des Paronym-Projekts* della IDS), WP (*Wikipedia Artikel u. Artikel- & Benutzerdiskussionen 2013/15/17/19*), TAGGED-C (*Archiv morphosyntakt. annotierter Korpora*) eccetera.

22 Per informazioni sugli archivi cfr. <https://www2.ids-mannheim.de/cosmas2/projekt/referenz/archive.html> [21.10.2022].

23 Per esempio, *W-öffentlich* (*alle öffentlichen Korpora des Archivs W mit Neuakquisitionen*); *lit-pub* (*Belletristik/Trivialliteratur öffentlich*); *focus* (FOCUS 2000 - 2021). Per informazioni sui corpora cfr. <https://www2.ids-mannheim.de/cosmas2/projekt/referenz/korpora.html> [21.10.2022].

24 Per informazioni sui corpora virtuali cfr. <https://www2.ids-mannheim.de/cosmas2/projekt/referenz/virtuell.html> [21.10.2022].

di avere un quadro chiaro delle varie possibilità. Cliccando su *Suchen* si dà avvio alla ricerca;

(4) Il primo risultato della ricerca è una lista di forme di parola (*Wortformen*), disposte in ordine alfabetico ascendente, dell'unità linguistica cercata. Per esempio, se si cerca *Nachhaltigkeit*, viene prodotta la seguente lista (Figura 3.3):

Liste zu: Nachhaltigkeit : 6 Wortformen, alphab. aufsteigend	
✓ NACHHALTIGKEIT	229
✓ NAchhAltigkeit	1
✓ NACHhaltigkeit	3
✓ NaCHHALTIGKEIT	1
✓ Nachhaltigkeit	107.794
✓ nachhaltigkeit	91

**Figura 3.3:** Screenshot della lista di forme parole prodotte dalla ricerca di *Nachhaltigkeit* in COSMAS II

Come si evince dai dati presenti in questa lista<sup>25</sup> (Figura 3.3), molte di queste forme di parola sono attribuibili a errori di scrittura. Eliminando la spunta verde accanto alla forma di parola che non si vuole analizzare (sulla parte sinistra della figura), tale forma di parola non verrà considerata nel prosieguo dell'analisi;

(5) una volta scelte le forme idonee alla propria ricerca, cliccando su *Ergebnisse* (risultati) verranno prodotti i risultati della ricerca<sup>26</sup> (*Treffer*) divisi per corpus (*Korpusansicht*): 108.119 risultati in 77.513 testi, risalenti all'arco temporale 1955-2021, provenienti da 833 corpora. È possibile poi visualizzare a quali corpora appartengono i risultati della ricerca<sup>27</sup>. La visualizzazione per corpus (*Korpusansicht*) è impostata di default dal sistema; i risultati possono però anche essere visualizzati utilizzando criteri diversi: fonte (*Quellenansicht*), documento (*Dokumentansicht*), paese di provenienza (*Ansicht nach Ländern*), tipologia testuale (*Ansicht nach Textsorten*), tema (*Ansicht nach Themen*), decennio (*Ansicht nach Jahrzehnt*), anno (*Jahr*), mese (*Monat*), giorno (*Tag*);

(6) Successivamente, cliccando sull'analisi delle co-occorrenze (*Ko-Occurrenzenanalyse*)<sup>28</sup>, è possibile procedere ad analisi di tipo sintagmatico (Figura 3.4):

25 NACHHALTIGKEIT – 229 occorrenze; NAchhAltigkeit – 1 occorrenza; NACHhaltigkeit – 3 occorrenze; NaCHHALTIGKEIT – 1 occorrenza; Nachhaltigkeit – 107.794 occorrenze; nachhaltigkeit – 91 occorrenze.

26 Nel caso esemplificato di *Nachhaltigkeit* non è stata eliminata nessuna forma di parola.

27 Nel caso esemplificato: 833 corpora visualizzati nel numero di 20 corpora per pagina, per un totale di 42 pagine.

28 Un altro strumento realizzato dal *Leibniz-Institut für Deutsche Sprache* per l'analisi delle co-occorrenze è la *Ko-Occurrenzenbank* (CCDB), che permette la visualizzazione e l'analisi di 220.309 profili di parola automatici. La CCDB permette, inoltre, di determinare le parole che hanno un

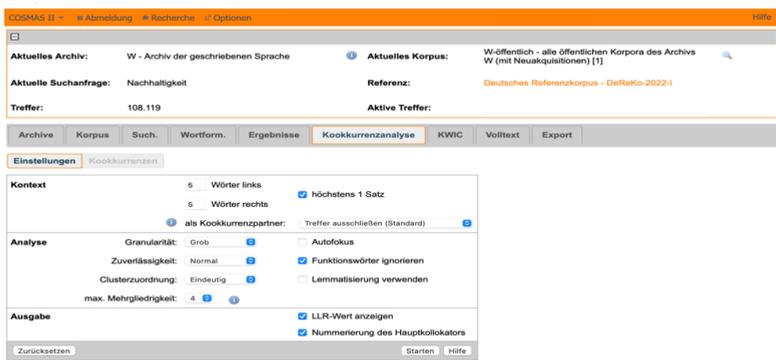


Figura 3.4: Screenshot della maschera di ricerca per le co-occorrenze (*Kookkurrenzanalyse*) in COSMAS II

Lasciando inalterati i valori impostati di default (5 parole a sinistra, 5 parole a destra per frase) i risultati, ordinati per il parametro statistico *Log-Likelihood-Ratio* (LLR)<sup>29</sup>, con indicazione anche della frequenza assoluta, mostrano le parole che co-occorrono in maniera più significativa con la parola ricercata (*Kookkurrenzen*) e i loro pattern sintagmatici più frequenti (*syntagmatische Muster*) (Figura 3.5):

#	LLR	kumul.	Häufig	Kookkurrenzen	syntagmatische Muster
1	22974	6021	6021	Thema	89% zum/das Thema [...] Nachhaltigkeit 10% Nachhaltigkeit ist ein [...] Thema
2	11594	7492	1471	ökologische	85% und ökologische (und ...) Nachhaltigkeit 14% der Nachhaltigkeit [...] und [...] ökologische ... und
3	11048	8732	1240	Ökologie	98% Ökologie (und) Nachhaltigkeit 41% Nachhaltigkeit (und) Ökologie und
4	10395	10075	1343	Umweltschutz	53% Umweltschutz (und) Nachhaltigkeit 46% Nachhaltigkeit (und) Umweltschutz
5	7983	11208	1133	Klimaschutz	51% Klimaschutz (und) Nachhaltigkeit 48% Nachhaltigkeit (und) Klimaschutz

Figura 3.5: Screenshot dei risultati della *Kookkurrenzanalyse* per *Nachhaltigkeit* in COSMAS II

profilo collocazionale simile alla parola cercata (*related collocation profiles*), di distinguere due parole in base al profilo collocazionale (*contrast near-synonyms*), di focalizzare aspetti quali il *topographic profile* con la *Self-Organizing Map* (SOM) o di individuare parole simili da un punto di vista semantico (*semantic proximity model*). Dato che l'ultimo aggiornamento risale al 2007, la CCDB non è adatta a individuare le tendenze più attuali della lingua tedesca. Per sopperire a ciò è stato realizzato DEREKOVECS (cfr. Fankhauser/Kupietz 2019; Kupietz/Lüngen/Kamocki/Witt 2018). Cfr. <http://corpora.ids-mannheim.de/openlab/derekovecs/> [21.10.2022].

29 Per informazioni sul parametro, cfr. Perkuhn/Keibel/Kupietz (2012: 118). Per un confronto tra i vari parametri (*chi-square*, *mutual information*, *log-likelihood ratio test*, *minimum sensivity*, *fisher's exact test*) per l'individuazione di collocazioni, si veda Stefanowitsch (2020: 224-233).

Per ogni risultato possono essere visualizzate le KWIC o le porzioni di testo con relativa indicazione delle fonti;

(7) In seguito, possono essere anche estratte le KWIC dell'unità linguistica cercata. Per default esse sono visualizzate in modalità non organizzata (*un-sortiert*), ma vi sono anche ulteriori possibilità: in ordine alfabetico (*alphabetisch sortiert*), in ordine cronologico (*chronologisch sortiert*), in ordine casuale (*zufällig sortiert*), in ordine di relazione di co-occorrenza (*Ko-okkurrenzstärke*). Oltre alle KWIC, cliccando su testo completo (*Volltext*) è possibile visualizzare i risultati con una maggiore quantità di testo. Scegliendo la funzione *Export* i dati, con un limite di occorrenze preimpostato, possono essere esportati in vari formati e analizzati sul proprio dispositivo.

L'uso di COSMAS II è abbastanza intuitivo (cfr. Lemnitzer 2022: 414) per un livello di analisi di base, mentre, per ricerche più approfondite, e anche più complesse da un punto di vista morfologico, è necessario acquisire una maggiore *corpus literacy* (cfr. anche Steyer 2008: 189).

### 3.2.2 Il DWDS – *Digitales Wörterbuch der Deutschen Sprache*

La piattaforma DWDS (*Digitales Wörterbuch der Deutschen Sprache*) della *Berlin-Brandenburgische Akademie der Wissenschaften* contiene numerosi corpora creati originariamente per fini lessicografici, ma utilizzabili oggi per analisi di tipo linguistico (Lemnitzer 2022: 414). Questo sistema lessicale digitale comprende corpora (*Textkorpora*), dizionari (*Wörterbücher*)<sup>30</sup> e anche strumenti per l'estrazione di dati statistici (*Wortstatistiken*).

I corpora, tutti monolingui in lingua tedesca, sono raggruppati in quattro gruppi: corpora di riferimento (*Referenzkorpora*), corpora di testi giornalistici (*Zeitungskorpora*), *web corpora* (*Webkorpora*) e corpora specialistici (*Spezialkorpora*). Sono prevalentemente corpora di lingua scritta, ma si possono trovare anche corpora di lingua parlata (per esempio, il *Wendekorpus*, il corpus *Text+Berg*, il corpus *Politische Reden* (1982-2020)) e corpora CMC come il corpus *Blogs*.

Tra i corpora di riferimento, che sono bilanciati sia per la dimensione diacronica, sia per i generi testuali, occorre annoverare tre corpora: il DWDS-*Kernkorpus* (1900-1999), il DWDS-*Kernkorpus 21* (2000-2010) e il *Deutsches Textarchiv*

30 Il dizionario principale è il DWDS-*Wörterbuch*, che mette a disposizione informazioni sulla forma e sul significato di ben 230.000 parole. Vi sono, inoltre, altre risorse lessicografiche: *Das Etymologische Wörterbuch des Deutschen* di Wolfgang Pfeifer, *OpenTheaurus*, *Das Wörterbuch der deutschen Gegenwartssprache* (WDG) e *Das Deutsche Wörterbuch von Jacob Grimm und Wilhelm Grimm* (DWB). Le informazioni presenti nel DWDS-*Wörterbuch*, in *Das Etymologische Wörterbuch des Deutschen* e in *OpenTheaurus* vengono mostrate anche durante la ricerca nel sistema informativo lessicale DWDS. Per esempio, si possono consultare informazioni quali la frequenza di una parola nel corpus, le informazioni grammaticali, la pronuncia, la sillabazione, le informazioni ortografiche (*Rechtschreibung*), le informazioni relative al significato con esempi d'uso, sinonimi, la *Wortverlaufskurve*, le attestazioni nei dizionari e nei corpora.

(1598-1913). Il corpus più grande è il *Deutsches Textarchiv* con 151.546.384 token (al 22.10.2022).

Tra i corpora giornalistici sono disponibili i principali giornali tedeschi sia digitali sia retrodigitalizzati<sup>31</sup> come, per esempio, *Der Tagesspiegel* e *Die Zeit* (1946-2018).

I *web corpora* comprendono sia corpora generici come il *Webkorpus*, sia corpora parziali (*Teilkorpora*) con focus su particolari tematiche, come la moda (cfr. *Mode und Beauty-Blogs*), la pandemia dovuta al Coronavirus (cfr. il *Corona-Korpus*). Aggiornato quotidianamente e molto interessante per studi che indagano nuove parole è il corpus *Webmonitor*, per il quale è però necessaria una registrazione.

La piattaforma ospita anche corpora specialistici, come il corpus *Polytechnisches Journal*, il corpus *Filmuntertitel*, il corpus *Briefe von Jean Paul* (1780-1825).

Tutti i corpora sono annotati automaticamente<sup>32</sup> e sono da considerarsi corpora dinamici. Molti dei corpora presenti sulla piattaforma sono fruibili liberamente, alcuni dei quali previa registrazione. Tutti i corpora forniscono metadati con informazioni su genere testuale, autore, titolo, casa editrice e data di pubblicazione.

La piattaforma DWDS offre diversi strumenti per la ricerca:

(1) il motore di ricerca linguistico DDC, che permette di effettuare una ricerca<sup>33</sup> per parole coniugate, parole alla forma base (lemmi), parti del discorso in vari corpora (cfr. Lemnitzer/Diewald 2022: 379) (Figura 3.6).

Korpora im DWDS

Korpusabfrage

Suche in Korpora

Korpus: DWDS-Kernkorpus (1900-19) Start: 1900 Ende: 1999 Textklassen: Belletristik Wissenschaft Gebrauchsliteratur Zeitung

Anzeige: KWIC voll maximal Sortierung: Datum absteigend Treffer pro Seite: 10

Übersicht über die Korpora im DWDS

Hinweis: Weitere Statistiken finden Sie auf einer separaten Seite.

Korpus	Kategorie	Zugang*	Tools	Dokumente	Sätze	Tokens
DWDS-Kernkorpus (1900-1999)	Referenzkorpora	frei	bc	79190	5821108	121494429
DWDS-Kernkorpus 21 (2000-2010)	Referenzkorpora	frei	bc	12184	874113	15469000
DTA-Kernkorpus (1598-1913)	Referenzkorpora	frei	l	1473	6179679	151546384
DTA-Kern+Erweit. (1465-1969)	Metakorpora	frei	l, bc	5427	10664817	246960570
Historische Korpora (1465-1998)	Metakorpora	frei	l, bc	64092	71454366	1484562129
Referenz- und Zeitungskorpora (frei)	Metakorpora	frei	v, bc	2469236	70343558	1330928502
ZDL-Regionalkorpus (ab 1993)	Metakorpora	Anmeldung	v, bc	3202249	638830786	9303084491
WebXL	Metakorpora	Anmeldung	v, bc	25754616	779055652	11419304555
Berliner Zeitung (1945-1993)	Zeitungskorpora	Anmeldung	bc, f	1570943	29230087	497848551
Berliner Zeitung (1994-2005)	Zeitungskorpora	frei	bc	851194	14575324	237011121

Figura 3.6: Screenshot della maschera di ricerca nel sistema DWDS

31 I testi retrodigitalizzati sono testi originariamente a stampa, cartacci, trasformati in formato digitale.

32 È possibile consultare anche la versione grezza di un corpus, ma questa opzione è prevista solo per un numero limitato di corpora. Cfr. <https://www.dwds.de/d/plot> [25.10.22].

33 Per le informazioni sulle possibilità di ricerca usando le espressioni regolari (*reguläre Ausdrücke*), arricchite da esempi concreti, cfr. <https://www.dwds.de/d/korpussuche> [25.10.22].

Come si evince dalla Figura 3.6, è possibile scegliere il corpus, l'arco temporale e il tipo di testo: testi letterari (*Belletristik*)<sup>34</sup>, testi scientifici (*Wissenschaft*), manualistica e saggistica (*Gebrauchsliteratur*)<sup>35</sup>, testi giornalistici (*Zeitung*). Ogni genere è collegato a un colore (rosso per testi letterari, giallo per testi scientifici, verde per manualistica e saggistica e blu per testi giornalistici). Inoltre, è possibile impostare la visualizzazione dei risultati per KWIC, per porzione di testo completa (*voll*), ossia la frase o per porzione di testo massima (*maximal*), ossia più frasi. Si può anche indicare la modalità di visualizzazione (*Sortierung*) e il numero preferito di risultati per pagina (*Treffer pro Seite*).

Una volta avviata la ricerca, oltre ai risultati, dai quali si identifica facilmente anche il genere di appartenenza tramite il colore, sul lato destro dello schermo si ha anche l'indicazione dei corpora dai quali sono estratti i dati (Figura 3.7):

**Korpusbelege DWDS-Kernkorpus (1900–1999)**

Nachhaltigkeit

Korpus: DWDS-Kernkorpus 1900-1999

Start: 1900 Ende: 1999

Textklassen: Belletristik Wissenschaft Gebrauchsliteratur Zeitung

Anzeige: KWIC voll maximal

Sortierung: Datum absteigend

Treffer pro Seite: 10

1-10 von 26 Treffern (27 insgesamt) Treffer exportieren

1. Archiv der Gegenwart, Bd. 67, 30.11.1997  
Die Beratungen des Treffens konzentrierten sich auf die demokratische Regierbarkeit, die Rechte der Kinder und Jugendlichen und die soziale Eingliederung, die Verbindung von sozialem Wissen und sozialer Entwicklung sowie auf die in diesem Zusammenhang stehenden Faktoren, die eine Entwicklung und die Nachhaltigkeit der gegenwärtigen Entwicklung in Lateinamerika erschweren.
2. G. A. [sAlt]: Nachhaltigkeit. In: Aktuelles Lexikon 1974-2000, München: DIZ 2000 [1997]  
Nachhaltigkeit (englisch: sustainability), dieser Begriff, der nächste Woche bei der Nachfolge-Konferenz des Rio-Gipfels wieder eine große Rolle spielen wird, kommt ursprünglich aus der Forstwirtschaft und bedeutet, daß man in einem Wald nicht mehr Bäume schlägt, als man wieder aufzueuten kann.
3. G. A. [sAlt]: Nachhaltigkeit. In: Aktuelles Lexikon 1974-2000, München: DIZ 2000 [1997]  
Daß das globale Ökosystem aus den Fugen geraten ist, führen Vertreter der Nachhaltigkeit auf einseitigen Raubbau an den Ressourcen der Erde zurück.
4. Die Zeit, 08.11.1996, Nr. 46  
Diese Schule hatte den heute geläufigen Grundsatz der "Nachhaltigkeit" vorgedacht.
5. Die Zeit, 08.11.1996, Nr. 46  
Doch der Begriff der "Nachhaltigkeit" drohte im Streit der Experten unterzugehen - die einen leiten aus ihm die Forderung ab, mit dem Abbau von Rohstoffen ganz aufzuhören, die anderen glauben, daß Rohstoffe im Prinzip durch Kapital ersetzbar sind.
6. Die Zeit, 08.11.1996, Nr. 46  
Gleichzeitig wurde Nachhaltigkeit zum Allweltswort auf politischen Konferenzen und Jahreshauptversammlungen von Großunternehmen - die ökologischen Ökonomen drohten zu Stichwortgebern für Werbelroschüren zu werden.
7. Die Zeit, 08.11.1996, Nr. 46  
Um dagegenzuhalten, arbeitet Nutzinger bis heute vor allem an der präzisen Formulierung von Regeln für die Nachhaltigkeit.
8. Die Zeit, 08.11.1996, Nr. 46  
So stellt sie die traditionelle Trennung von der Verteilung der Ressourcen und der Einkommen in Frage - denn Nachhaltigkeit läuft letztlich hinaus auf einen Interessensausgleich zwischen Erster und Dritter Welt und zwischen heutigen und künftigen Generationen.
9. G. A. Lexikon der Kunst - F. im. Olibrich, Harald (Hg.), Lexikon der Kunst, Berlin: Directmedia Publ. 2003 [1989], S. 9323  
Arbeit von großer Nachhaltigkeit.
10. Matthäus, Wolfgang: Carl: Die Musik in Geschichte und Gegenwart, Berlin: Directmedia Publ. 2002 [1979], S. 11253  
In Nachhaltigkeit und Steuerung dieser Rezeption ist sein Werk dem der bedeutendsten Meister seiner Zeit zur Seite zu stellen.

1-10 von 26 Treffern (27 insgesamt) Treffer exportieren

**Belege in Korpora**

Referenzkorpora  
DWDS-Kernkorpus (1900-1999) (27)  
DWDS-Kernkorpus 21 (2000-2010) (40)  
DTA-Kernkorpus (1598-1913) (48)

Manuskorpus  
DTA-Kern-Erwerb (1845-1869) (63)  
Historische Korpora (1465-1908) (279)  
Referenz- und Zeitungskorpora (frei) (232)

Zeitungskorpora  
Berliner Zeitung (1994-2005) (378)  
Der Tagesspiegel (ab 1994) (505)  
Die ZEIT (1946-2018) (2260)

Wikiskorpus  
Blogg (1134)

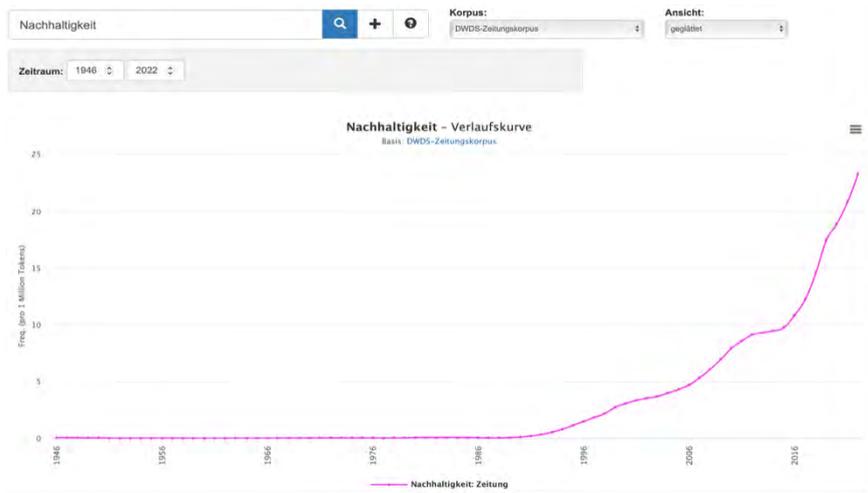
Spezialkorpora  
DTA-Erweiterungen (1405-1909) (15)  
Archiv der Gegenwart (1911-2000) (25)  
Polytechnisches Journal (18)  
Filmsamml (23)  
Gesprochene Sprache (5)  
DDR (4)  
Politische Reden (1962-2020) (1867)  
A. v. Humboldts Publizistik (dt., 1790-1809) (1)  
Nachrichten aus der Brüdergemeine (1819-1894) (0)  
Der Neue Pflanz (1842-1895) (1)  
Briefe von Jean Paul (1780-1825) (0)  
Wikipedia-Korpus (3876)

Figura 3.7: Screenshot dei risultati della ricerca di *Nachhaltigkeit* nel sistema DWDS

(2) la *Wortverlaufskurve* (curva di andamento della parola) con la quale si può analizzare l'andamento di un'unità linguistica in un certo arco temporale (Figura 3.8):

34 Cfr. <https://www.dwds.de/wb/Belletristik> [20.09.23].

35 Cfr. <https://www.dwds.de/wb/Gebrauchsliteratur> [20.09.23].



**Figura 3.8:** Screenshot dei risultati della ricerca di *Nachhaltigkeit* nella *Wortverlaufskurve* (DWDS)

La frequenza di una parola può, infatti, variare con l'andamento del tempo: può aumentare, come nel caso della parola *Nachhaltigkeit*, o diminuire; un'unità lessicale può addirittura andare in disuso (*Rechner*) o sostituirla con un'altra (*Sneakers* per *Turnschuhe*). È possibile anche visualizzare due unità linguistiche contemporaneamente e quindi confrontarle.

La *Wortverlaufskurve* copre un arco temporale di circa 400 anni; la ricerca delle parole avviene nei corpora DTA-Gesamt+DWDS-Kernkorpus (1600-1999, circa 368 milioni di token), nel DWDS-Zeitungskorpus (dal 1946, circa 22 miliardi di token) e nel ZDL-Regionalkorpus (dal 1993, circa 9,2 miliardi di token). Sono possibili diversi tipi di visualizzazione, in base alla frequenza normalizzata per un milione di token, in base alla frequenza assoluta.

(3) il *Wortprofil* (profilo di parola) per l'estrazione di profili di parole co-occorrenti<sup>36</sup> anche con focus su relazioni di tipo sintattico (cfr. anche Didakowski/Geyken 2014). Questo strumento è molto importante per determinare combinazioni di parole più o meno fisse, come le collocazioni<sup>37</sup>. La prospettiva di analisi è di tipo sincronico, a differenza dello strumento (4), che è di tipo diacronico.

Nella Figura 3.9 si mostra il risultato della ricerca del profilo di parola di *Nachhaltigkeit* con l'opzione "mostrami le prime 5 combinazioni più tipiche" in formato tabellare, mentre nella Figura 3.10 in formato *word cloud* (*Wortwolke*):

<sup>36</sup> Le collocazioni possono essere estratte in base alla loro frequenza di co-occorrenza nel corpus (*Frequenz*) o in base al *logDice* (cfr. Stefanowitsch 2020).

<sup>37</sup> Per la/le definizione/i di collocazione si veda il paragrafo 2.6.

Überblick $\mathbb{A}, \mathbb{B}$			hat Adjektivattribut $\mathbb{A}, \mathbb{B}$			ist Akk./Dativ-Objekt von $\mathbb{A}, \mathbb{B}$		
	logDice $\uparrow$	Freq. $\uparrow$		logDice $\uparrow$	Freq. $\uparrow$		logDice $\uparrow$	Freq. $\uparrow$
1. Prinzip	8.2	517	1. ökologisch	7.8	710	1. achten	6.9	164
2. Kriterium	8.2	146	2. offensive	7.3	9	2. zweifeln	6.6	93
3. ökologisch	7.8	710	3. fiskalisch	6.1	8	3. verschreiben	5.2	29
4. Aspekt	7.8	181	4. finanzpolitisch	4.8	10	4. gewährleisten	4.5	32
5. Gedanke	7.6	145	5. unternehmerisch	4.8	24	5. definieren	4.2	21

ist in Präpositionalgruppe $\mathbb{A}, \mathbb{B}$			hat Genitivattribut $\mathbb{A}, \mathbb{B}$			hat in Koordination mit $\mathbb{A}, \mathbb{B}$		
	logDice $\uparrow$	Freq. $\uparrow$		logDice $\uparrow$	Freq. $\uparrow$		logDice $\uparrow$	Freq. $\uparrow$
1. Zweifel an	7.4	242	1. Spa-bereich	6.5	9	1. und Klimaschutz	10.0	126
2. ausrichten auf	6.6	69	2. Übergabeprozess	6.1	7	2. und Umweltschutz	9.5	95
3. Bewußtsein für	6.3	43	3. Staatsfinanzen	6.1	27	3. und Generationengerechtigkeit	9.0	32
4. Duftelson von	5.9	27	4. Konjunkturerholung	6.1	13	4. & Leben	8.5	23
5. Fokus auf	5.7	22	5. Konjunkturaufschwung	5.6	8	4. und Phantasie	8.4	19

hat Prädikativ $\mathbb{A}, \mathbb{B}$			ist Subjekt von $\mathbb{A}, \mathbb{B}$			hat Präpositionalgruppe $\mathbb{A}, \mathbb{B}$		
	logDice $\uparrow$	Freq. $\uparrow$		logDice $\uparrow$	Freq. $\uparrow$		logDice $\uparrow$	Freq. $\uparrow$
1. fraglich	2.9	11	1. bedauern	5.5	230	1. in Finanzierung	7.9	28
2. Stichwort	2.8	5	2. draufstehen	3.8	6	2. in Immobilienwirtschaft	5.9	5
3. Thema	2.5	214	3. spielen	3.4	144	3. in Landwirtschaft	5.2	13
4. Begriff	2.5	45	4. vertragen	3.2	5	4. im Bau	5.0	5
5. Kriterium	2.3	18	5. definieren	2.9	5	5. in Mode	4.9	8

ist Genitivattribut von $\mathbb{A}, \mathbb{B}$			ist Prädikativ von $\mathbb{A}, \mathbb{B}$			ist in vergl. Wortgruppe $\mathbb{A}, \mathbb{B}$		
	logDice $\uparrow$	Freq. $\uparrow$		logDice $\uparrow$	Freq. $\uparrow$		logDice $\uparrow$	Freq. $\uparrow$
1. Prinzip	8.2	517	1. engagieren	4.3	18	1. Thema wie	7.6	129
2. Kriterium	8.2	146	2. einsetzen	4.2	54	2. Wert wie	6.0	33
3. Aspekt	7.8	181	3. stehen	3.8	75	3. Begriff wie	5.2	16
4. Gedanke	7.6	145	4. tun	3.0	12	4. Kriterium wie	3.6	8
5. Begriff	7.3	217	5. werben	2.9	14			

Figura 3.9: Screenshot dei risultati della ricerca di *Nachhaltigkeit* nel *Wortprofil* (DWDS).  
Formato tabella

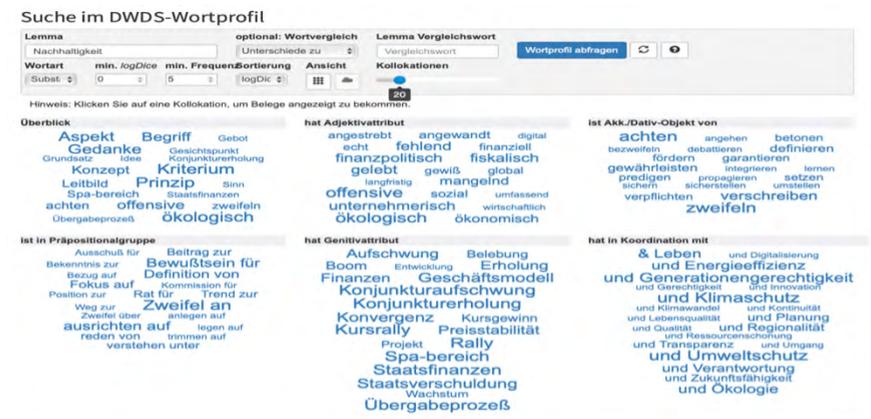


Figura 3.10: Screenshot dei risultati della ricerca di *Nachhaltigkeit* nel *Wortprofil* (DWDS).  
Formato *word cloud*

Ad eccezione della categoria generale (*Überblick*), in cui non vi è differenziazione delle relazioni sintattiche, negli altri riquadri i risultati sono organizzati per relazioni sintattiche: vi sono, infatti, informazioni su quando *Nachhaltigkeit* è soggetto (*Subjekt*), oggetto diretto (*Akkusativobjekt*), oggetto indiretto (*Dativobjekt*), attributo al genitivo (*Genitivattribut*), aggettivo in funzione attributiva (*Adjektivattribut*) eccetera. Tramite il *Wortprofil* si possono identificare anche combinazioni lessicali di più parole (*Mehrwortausdrücke*) oppure si possono mettere a confronto due profili di parola.

(4) *DiaCollo* per l'estrazione di co-occorrenze in un preciso arco temporale. La prospettiva adottata è di tipo diacronico. Con questo strumento è possibile

analizzare la variazione delle co-occorrenze di una unità linguistica e le loro frequenze nel corso del tempo (Figura 3.11).

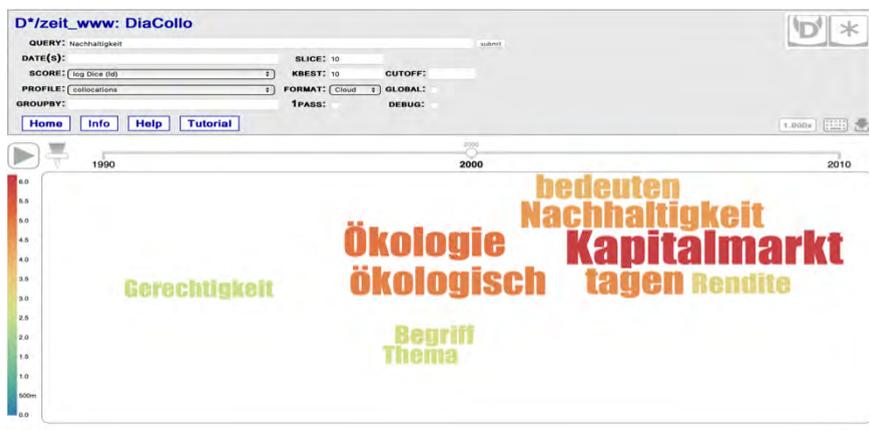


Figura 3.11: Screenshot dei risultati della ricerca di *Nachhaltigkeit* in *DiaCollo* (DWDS)

È possibile visualizzare i risultati tramite il formato *Word-Cloud*<sup>38</sup> (*Wortwolke*), secondo cui le parole più grandi sono quelle che co-occorrono più frequentemente (in questo caso *Kapitalmarkt*). Anche i colori svolgono un ruolo importante in questa rappresentazione grafica: le parole evidenziate nello stesso colore presentano anche il medesimo punteggio (*score*). Cliccando su una parola del *cloud* appaiono non solo informazioni numeriche importanti (la parte del discorso della parola, l'anno, lo *score*, i dettagli sulla frequenza ecc.), ma anche il link alle KWIC. Cliccando invece sul tasto *play*, si può avere una dimostrazione animata dell'andamento delle collocazioni nell'arco temporale scelto.

Questo strumento di analisi non offre indicazioni sulle relazioni grammaticali tra le due parole co-occorrenti.

Il sistema DWDS è molto versatile (cfr. Steyer 2008: 190) e la ricerca nella DDC intuitiva (cfr. Lemnitzer 2022: 415). I risultati possono essere scaricati in vari formati. Le sue applicazioni, anche con finalità didattiche, sono state mostrate in numerosi saggi (cfr. Wallner 2013; Flinz/Katelhön 2019; Krekeler 2021; Malloggi 2021).

### 3.2.3 La *Datenbank für Gesprochenes Deutsch* (DGD)

La *Datenbank für Gesprochenes Deutsch* (DGD)<sup>39</sup> è una raccolta di corpora di lingua parlata messa a disposizione dal *Leibniz-Institut für Deutsche Sprache* per

38 Altre possibilità di visualizzazione sono *Highchart* e *Bubble*.

39 Cfr. <https://dgd.ids-mannheim.de/> [24.10.22].

ricerche di tipo linguistico che include anche un software di analisi. Alla DGD si accede liberamente previa registrazione.

Con la DGD si possono consultare i metadati dei corpora, delle registrazioni e dei soggetti parlanti, i dati audio, le trascrizioni, i video e gli eventuali materiali aggiuntivi (informazioni sul setting e svolgimento delle interazioni, liste di frequenza di parole e di lemmi, convenzioni per la trascrizione ecc.). Questi dati sono collegati tra loro: per esempio, dalla trascrizione si può accedere al dato audio e ai metadati del parlante. È possibile fare ricerche semplici nel testo (*Volltextsuche*), ricerche esplorative di forme linguistiche, ma anche ricerche più specifiche e complesse (cfr. Deppermann/Schmidt 2014). I risultati della ricerca (KWIC) possono essere raffinati qualitativamente tramite filtri e si possono costruire corpora virtuali. I dati possono poi essere esportati in vari formati.

La DGD comprende ben 40 corpora di lingua tedesca parlata di diverse tipologie: corpora di interviste biografiche (il *Berliner Wendekorpus*, il “cosiddetto” *Israelkorpus* con due corpora di prima generazione, IS – *Emigrantendeutsch in Israel*<sup>40</sup> e ISW – *Emigrantendeutsch in Israel: Wiener in Jerusalem*<sup>41</sup> e uno di seconda generazione ISZ – *Zweite Generation deutschsprachiger Migranten in Israel*<sup>42</sup>, il corpus FEGB – *Flucht und Emigration nach Großbritannien*<sup>43</sup>, il corpus FGOP – *Fluchtgeschichten aus Ostpreußen*<sup>44</sup> ecc.), corpora specialistici (il corpus GeWiss – *Gesprochene Wissenschaftssprache*<sup>45</sup>), corpora video (il corpus *Belgische TV-Debatten*<sup>46</sup>), corpora per l’analisi della variazione diatopica (il corpus *Deutsch Heute*<sup>47</sup> ecc.) anche extraterritoriale (il corpus *Australiendeutsch*<sup>48</sup> ecc.), il corpus di riferimento per la lingua parlata FOLK – *Forschungs- und Lehrkorpus Gesprochenes Deutsch* (cfr. Deppermann/Schmidt 2014). Non tutti i corpora presenti nella DGD sono annotati linguisticamente.

FOLK<sup>49</sup>, nella sua versione 2.20 del 15.06.2023, consta di 414 registrazioni (*Ereignisse*) con 1.317 parlanti. Per 142 di queste sono presenti anche file video.

40 Cfr. <http://hdl.handle.net/10932/00-0332-C3A7-393A-8A01-3>. Cfr. Betten (1995); Betten/Flinz/Leonardi (2023). Per le pubblicazioni sull’*Israelkorpus* cfr. <https://www.zotero.org/groups/2219390/israelkorpus/library> [24.10.22].

41 Cfr. <http://hdl.handle.net/10932/00-0332-C42A-423C-2401-D> [24.10.22].

42 Cfr. <http://hdl.handle.net/10932/00-0332-C453-CEDC-B601-2> [24.10.22].

43 Cfr. <https://dgd.ids-mannheim.de/> [24.10.22]. Al momento della pubblicazione non è ancora stato attribuito un PID (*Persistent Identifier*). Cfr. Thüne (2019).

44 Cfr. <https://dgd.ids-mannheim.de/> [24.10.22]. Al momento della pubblicazione non è ancora stato attribuito un PID (*Persistent Identifier*). Cfr. Cinato (2020).

45 Cfr. <http://hdl.handle.net/10932/00-0332-BCF9-BE93-5F01-E> [24.10.22]. Cfr. Fandrych/Meißner/Slavcheva (2012).

46 Cfr. <http://hdl.handle.net/10932/00-03FA-9CB0-5E33-8E01-8> [24.10.22].

47 Cfr. <http://hdl.handle.net/10932/00-0439-7224-E314-7301-1> [24.10.22]. Cfr. Brinckmann/Kleiner/Knöbl/Berend (2008).

48 Cfr. <http://hdl.handle.net/10932/00-0332-BCF9-BE93-5F01-E> [24.10.22]. Cfr. Clyne (1981).

49 Cfr. <http://hdl.handle.net/10932/00-0332-C1B2-A5E3-2A01-D> [24.10.22]. Cfr. Reineke/Deppermann/Schmidt (2023).

La durata delle registrazioni è di 347 ore e 2 minuti e le trascrizioni sono 883. Si tratta di un corpus dinamico, che viene aggiornato annualmente con un incremento di 30 ore di registrazioni annue.

Obiettivo primario del corpus FOLK è quello di mettere a disposizione una varietà bilanciata di conversazioni autentiche e spontanee per rappresentare l'interazione linguistica nelle sue varie sfaccettature. Il corpus si presta per analisi di tipo linguistico di vario tipo (cfr. anche Deppermann/Hartung 2011; Kaiser 2018: 516). Le registrazioni coprono un ampio spettro di situazioni:

- private: per esempio conversazioni durante i pasti, durante le attività quotidiane come il cucinare o il passeggiare, durante il gioco; conversazioni telefoniche;
- istituzionali: a scuola e all'università (lezioni, ricevimenti), in luoghi pubblici come farmacie, negozi, scuole guida (compra-vendita, lezioni guida), in luoghi di lavoro (meeting, cambio turno in ospedale);
- pubbliche: dibattiti parlamentari, tavole rotonde.

È possibile isolare parametri quali la provenienza, l'età e la formazione dei soggetti parlanti. La trascrizione delle conversazioni è stata eseguita seguendo le convenzioni cGAT-*Minimaltranskript* (cfr. Schmidt/Schütte/Winterscheid 2015) con una trascrizione ortografica (*literarische Umschrift*) che mira a rappresentare graficamente fenomeni del parlato spontaneo come p.es. le legature (cfr. *baste* in Schmidt 2022b: 423). Altri livelli di annotazione sono: la normalizzazione ortografica, con la quale ogni parola trascritta viene uniformata alla variante ortografica standard, la lemmatizzazione, in cui ogni token viene assegnato alla sua forma base, e l'annotazione morfosintattica secondo il *Stuttgart-Tübingen-Tagset* (STTS).

Il corpus FOLK può essere analizzato sia da un punto di vista quantitativo, sia qualitativo. I dati possono anche essere esportati e quindi essere analizzati con altri strumenti per varie finalità, anche didattiche (cfr. Imo/Weidner 2018; Kaiser/Schedl 2021). Proprio per incrementare l'uso di FOLK nella didattica del tedesco come lingua straniera e lingua seconda (*DaF* e *DaZ*), è stato realizzato il progetto *ZuMult – Zugänge zu multimodalen Korpora gesprochener Sprache*<sup>50</sup>, che permette la scelta di audio per livello linguistico (cfr. Fandrych/Meißner/Wallner 2018, Fandrych/Schmidt/Wallner/Wörner 2023). La scelta degli eventi (*ZuMal*) può essere fatta in base a parametri quali il tipo di conversazione, il tema, la provenienza, il livello linguistico, la vicinanza al tedesco standard, la velocità di eloquio, i turni di parola sovrapposti tra parlanti.

Da FOLK si può accedere sia alle trascrizioni della DGD, sia a quelle di *ZuMult* (*Tool Search by Query*). Queste ultime sono state modellate appositamente per finalità didattiche, con colori per un supporto grafico, una barra

50 Cfr. <https://zumult.org> [24.10.22].

di navigazione per la visualizzazione del cambio di parlante e per il grado di interazione<sup>51</sup>.

### 3.2.4 Altri corpora

Altri corpora in cui è presente la lingua tedesca sono disponibili su CLARIN<sup>52</sup> o sul *Virtual Language Observatory*<sup>53</sup>, raggruppati per categorie (cfr. paragrafo 3.1)

Segue una breve descrizione di alcuni corpora:

(1) corpora per l'apprendimento e l'insegnamento (*Lernerkorpora*) della lingua tedesca: oltre ai corpora *Hamburg Maptask*<sup>54</sup>, *Berlin Maptask*<sup>55</sup>, *Hamburg Modern Times*<sup>56</sup>, che contengono dati di studentesse e studenti<sup>57</sup> (con un livello di tedesco avanzato) che svolgono un compito in cui si devono descrivere delle mappe (*maptask*), vi sono i corpora Falko e MERLIN. Il corpus Falko<sup>58</sup>, che può essere consultato tramite il tool ANNIS, contiene testi prodotti da apprendenti di tedesco con diverse L1 durante prove di esame (vari livelli linguistici), e comprende sei sotto-corpora che si distinguono per tipo di compito di scrittura e L1. I dati sono annotati per morfosintassi e per errori. Il corpus MERLIN – *Multilingual Platform for European Reference Levels: Interlanguage Exploration in Context*<sup>59</sup> raccoglie testi scritti prodotti da apprendenti di italiano, ceco e tedesco durante prove di esame (livelli A1-C1). Il corpus è annotato a più livelli;

(2) corpora per l'analisi della comunicazione in Internet<sup>60</sup> (*Internetbasierte Kommunikation, Computer-Mediated Communication*), che comprendono dati provenienti da comunicazioni dialogico-interattive (in Internet), per esempio da applicazioni chat (*WhatsApp*), da social media (*Facebook, Instagram*), da servizi di weblog o microblogging (*Twitter*), da tecnologie per conferenze (*Zoom, Google-Meet, Microsoft-Teams* ecc.), da portali video (*YouTube*), da piattaforme per l'apprendimento (*Moodle* ecc.), da giochi online (cfr. Beißwenger/Lüngen 2022: 431). Si tratta di corpora molto utili per l'analisi di pratiche discorsive tipiche

51 Cfr. <https://zumult.org/demo/> [24.10.22].

52 Cfr. <https://www.clarin.eu> [24.10.22].

53 Cfr. <https://vlo.clarin.eu/?1> [24.10.22].

54 Cfr. <https://www.fdr.uni-hamburg.de/record/1481> [03.1.23].

55 Cfr. <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/bematac/bematac#Zugang> [24.10.22].

56 Cfr. <https://www.fdr.uni-hamburg.de/record/1483> [03.1.23].

57 Per studi sull'apprendimento del tedesco da parte di bambini e bambine, si citano qui anche il corpus DULDE (*Deutscher und Französischer Doppelter Erstsprachenerwerb*), basato su uno studio longitudinale, e il corpus MEKI (*Mebrsprachige Kinder im Vorschulalter*), che analizza la competenza di apprendenti con background plurilingue ed è consultabile nella DGD. Cfr. Schmidt (2022b: 429).

58 Cfr. *Falko-Gesamtkorpus V2.0*. Cfr. <https://hu-berlin.de/falko> [24.10.22].

59 Cfr. <https://merlin-platform.eu/index.php> [23.10.22].

60 Per informazioni su altri corpora, come *SMS4science.ch, WhatsUp Switzerland, It-Blogs, Mode-und Beauty Blogs, DDI-Korpus* cfr. Beißwenger/Lüngen (2022: 435-439).

di questo tipo di comunicazione, caratterizzato da un nuovo genere di scrittura<sup>61</sup>. Il *Dortmunder Chat-Korpus* è un corpus che contiene dati presi da *Logfiles* da vari tipi di chat. Sviluppato dall'Università di Dortmund<sup>62</sup> per studiare la lingua della comunicazione in Internet, è stato successivamente ampliato e messo a disposizione sia tramite DeReKo, sia tramite il sistema informativo DWDS (cfr. Beißwenger/Storrer 2008; Beißwenger 2013). È annotato non solo da un punto di vista morfosintattico, ma anche per le caratteristiche tipiche della comunicazione in Internet (emoticons ecc.). Altri corpora sono:

- il corpus MoCoDa I (*Mobile Communication Database I*)<sup>63</sup> realizzato dall'Università di Münster, che comprende SMS e altri tipi di messaggi (1.063.502 token), nonché relativi metadati per studi anche micro-diacronici (cfr. Imo 2017: 95-98);
- il corpus MoCoDa 2 (*Mobile Communication Database 2*)<sup>64</sup>, realizzato dall'Università di Duisburg-Essen, con messaggi soprattutto da *WhatsApp* (299.681 token) annotati e comprensivi di metadati (cfr. Beißwenger/Imo/Fladrich/Ziegler 2019);
- i corpora comparabili *Wikipedia* in DeReKo. Per ogni lingua sono raccolti tutti gli articoli enciclopedici della *Enciclopedia libera Wikipedia*<sup>65</sup>, tutte le pagine di discussione, caratterizzate da un tipo di scrittura collaborativo, nonché tutte le pagine “discussioni utenti” (cfr. Längen/Kupietz 2017). I corpora *Wikipedia* sono consultabili non solo tramite COSMAS II, ma anche tramite KorAP, dove, oltre all'annotazione morfosintattica, sono dotati di annotazione sintattica (costituenti e dipendenza). Possono essere anche scaricati;
- i corpora *Blogs* del DWDS, con testi a partire dal 1990, e che constano di 109 milioni di token da 229.000 documenti blog. I dati sono stati estratti dalla piattaforma *WordPress.com* (cfr. Barbaresi/Würzner 2014);

(3) *web corpora* comparabili plurilingui. Nella piattaforma di *Sketch Engine* sono consultabili i corpora appartenenti alla famiglia *TenTen*, che sono corpora molto grandi (più di 10 bilioni di parole per lingua) dotati di metadati, lemmatizzati e annotati da un punto di vista morfosintattico<sup>66</sup>. Sono corpora statici, ma nel sistema sono disponibili più versioni: per il tedesco, per esempio, il *German Web 2013* (*deTenTen13*), il *German Web 2018* (*deTenTen18*), il *German Web*

61 Cfr. Per la descrizione di questa scrittura rivolta all'interazione (*interaktionsorientiertes Schreiben*) cfr. Storrer (2018). Per una panoramica sulla lingua tedesca nei social media cfr. Beißwenger/Längen (2022).

62 Cfr. <https://www.uni-due.de/germanistik/chatkorpus/> [25.10.22].

63 Cfr. <https://mocoda.spracheinteraktion.de> [25.10.22].

64 Cfr. <https://db.mocoda2.de/c/home> [25.10.22]. È prevista l'integrazione di MoCoDa 2 in DeReKo.

65 Cfr. [https://it.wikipedia.org/wiki/Pagina\\_principale](https://it.wikipedia.org/wiki/Pagina_principale) [25.10.22].

66 Cfr. <https://www.sketchengine.eu/documentation/tenten-corpora/> [25.10.22].

2020 (*deTenTen20*). I corpora non sono accessibili gratuitamente, ma solo a pagamento<sup>67</sup>;

(4) corpora plurilingui comparabili specialistici di lingua scritta, come il corpus *Lessico dei Beni Culturali* (LBC)<sup>68</sup>, che comprende testi specialistici (beni culturali) appartenenti a vari generi testuali in varie lingue (francese, inglese, italiano, russo, spagnolo, tedesco);

(5) corpora plurilingui comparabili specialistici di lingua parlata, come il corpus *Gesprochene Wissenschaftssprache Kontrastiv*<sup>69</sup> consultabile oggi tramite la DGD (cfr. paragrafo 3.2.4). Obiettivo primario del corpus GWSS è l'analisi di una particolare varietà di parlato tedesco, il linguaggio scientifico, con focus anche sul plurilinguismo. Consta di tre tipi di dati con l'obiettivo di rappresentare la comunicazione in ambito universitario: esami di laurea (*universitäre Prüfungsgespräche als abschließender Teil der Qualifikation zum Erwerb eines universitären Abschlusses*), relazioni da parte di esperti durante workshop o conferenze ecc., relazioni di studentesse e studenti in aula. Le registrazioni sono state fatte sia nei paesi coinvolti nel progetto (Inghilterra, Germania, Polonia), sia in altri paesi (Italia, Bulgaria, Finlandia). Nei paesi non di lingua tedesca sono state raccolte registrazioni anche di parlanti non madrelingua. GWSS permette, dunque, di operare confronti sulla base del tipo di conversazione (esame, relazione da parte di esperti, relazioni da parte di studentesse e studenti), della lingua (tedesco, inglese, polacco, italiano), della situazione linguistica (tedesco e inglese L1 vs. tedesco e inglese L2), del contesto accademico (Germania, Gran Britannia, Polonia, Bulgaria, Italia, Finlandia);

(6) il corpus parallelo di lingua scritta InterCorp, che contiene testi in varie lingue (tra cui italiano, francese, tedesco, spagnolo ecc.) allineati manualmente e appartenenti a varie tipologie: testi letterari, scientifici (92.290.000 parole) e di divulgazione. I subcorpora presenti in InterCorp sono dotati di metadati e alcuni corpora sono annotati dal punto di vista morfosintattico<sup>70</sup>.

67 Fino al 2022, grazie a un progetto europeo, *Sketch Engine* era fruibile gratuitamente previa registrazione. Oggi molte università offrono l'accesso a *Sketch Engine*; nel caso in cui si volesse fare una prova gratuita e la propria università non mettesse a disposizione l'accesso, l'account *free trial* di 30 giorni potrebbe essere una valida soluzione.

68 Cfr. <http://corpora.lesicobeniculturali.net/de> [25.10.22].

69 Inizialmente il corpus era consultabile sulla piattaforma dell'Università di Lipsia (Leipzig).

70 Cfr. <https://intercorp.korpus.cz/?lang=de> [25.10.22].



# Capitolo 4

## Costruire un corpus *ad hoc*

Il panorama dei corpora in lingua tedesca è molto ricco e variegato (cfr. paragrafo 3.2), e comprende corpora per soddisfare obiettivi e domande di ricerca di diverso tipo.

L'uso di corpora già preconfezionati è ampiamente raccomandato in letteratura, tanto che alla domanda «Selber kochen oder auswärts essen gehen?», dove “selber kochen” indica la costruzione di un corpus e “auswärts essen gehen” l'impiego di corpora già esistenti, Lemnitzer/Zinsmeister (2015: 136) rispondono «auswärts essen gehen» fornendo una serie di motivazioni condivisibili e riconducibili ai criteri sopracitati.

Per alcuni ambiti specialistici o per alcuni interessi di ricerca particolari, può però accadere che nessuno di questi corpora risulti indicato, o che lo sia solo parzialmente (cfr. Flinz 2018; Flinz 2021c). Pertanto, diventa necessario costruire un corpus *ad hoc*.

Nel presente capitolo è tematizzata proprio questa possibilità, partendo da alcune considerazioni preliminari circa la fase progettuale di creazione di un corpus fino alla fase di acquisizione dei dati (4.1). Nel paragrafo 4.2 sono presentati due possibili percorsi per la costruzione di un corpus per mezzo del tool *Sketch Engine* (4.2).

### 4.1 Considerazioni preliminari

La costruzione di un corpus è un procedimento molto complesso che può constare di varie fasi, tra cui:

- la progettazione: la fase progettuale è una fase molto importante, in cui si deve scegliere il tipo di corpus da costruire (cfr. paragrafo 3.1). Occorre dunque chiedersi quale tipo di corpus si adatti agli obiettivi della ricerca. Lemnitzer/Zinsmeister (2015: 137f), nella loro proposta di classificazione dei corpora, distinguono tra criteri che riguardano i dati primari (*Designkriterien*) e criteri utilizzabili per la preparazione di un corpus (l'annotazione);
- l'elaborazione di un modello per la costruzione del corpus: in questa fase si procede all'identificazione dei dati primari e alla definizione dei criteri alla base della loro selezione, riflettendo anche sulla quantità e sulla proporzione di testi per ogni tipologia individuata; inoltre, occorre decidere se utilizzare testi interi o frammenti di essi (tramite campionamento casuale). Il campionamento è, infatti, un'operazione necessaria (cfr. Freddi 2019: 12), che richiede una riflessione su alcune questioni fondamentali: Quanto

dovrà essere grande il campione? Quali criteri utilizzare per la selezione del materiale da includere nel campione (ossia l'intero progetto del corpus)? Sarà opportuno prediligere criteri esterni<sup>1</sup> all'uso linguistico come, per esempio, la variazione diamesica (asse scritto-parlato), la variazione diafasica (per genere, per registro), la variazione diastratica (sociale), diatopica (geografica)? Con i criteri esterni potranno essere fissate le molteplici dimensioni della variabilità intrinseca alla popolazione che il corpus vuole riprodurre per ottenere una rappresentazione quanto più bilanciata possibile della variabilità della popolazione (cfr. Freddi 2019: 13);

- la definizione della strutturazione del corpus: il corpus è strutturato in sottocorpora? Se sì, in base a quali criteri (macro-varietà identificate sulla base della macrostruttura e del paratesto)? Quali dimensioni? Quale è il loro rapporto?
- la definizione;
- la scelta del sistema informatico<sup>2</sup> in cui si inseriscono i dati primari;
- il reperimento e l'inserimento dei dati: sia nel caso di dati reperiti online, sia nel caso di dati presi da testi cartacei occorre sempre prevedere una procedura di pulizia dei testi e un controllo qualitativo;
- la tokenizzazione, lemmatizzazione e annotazione dei dati (cfr. Rossini Favretti 2000: 40).

## 4.2 Costruire un corpus *ad hoc* con *Sketch Engine*

Un corpus può essere costruito per mezzo di diversi sistemi informatici (tra cui *AntCon*, *Word Smith Tool*, *Corpus Explorer*).

Con *Sketch Engine* si può costruire un corpus scegliendo l'opzione *new corpus* dal *dashboard* (Figura 4.1) oppure dalla pagina *select corpus/advanced* (Figura 4.2):



**Figura 4.1:** Screenshot della parte superiore della pagina *dashboard* di *Sketch Engine*

- 1 Per ridurre al minimo l'intervento del ricercatore, dovranno essere privilegiati criteri esterni (cfr. Rossini Favretti 2000: 48).
- 2 Lemnitzer (2022a: 350-351) distingue tra *Korpusmanagementsystem* e *Korpusrecherchesystem*. Con il primo fa riferimento a un software che permette di amministrare una serie di corpora e di metterli a disposizione su una piattaforma; solitamente si possono anche salvare e amministrare i metadati, strutturare i dati primari e annotare i testi (cfr. Geyken et al. 2017). L'utente non viene a contatto con questo tipo di sistema. Con il secondo si intende un software che contiene non solo i corpora, ma che permette anche la loro consultazione tramite una serie di funzionalità. L'utente esegue le sue analisi con questo tipo di sistema.



Figura 4.2: Screenshot della parte superiore della pagina *select corpus* di *Sketch Engine*

Una volta scelta l'opzione *new corpus* devono essere inserite le informazioni relative al corpus che si andrà a creare:

- nome del corpus;
- tipo di corpus: monolingue o plurilingue;
- lingua dei dati primari;
- descrizione di alcune sue caratteristiche principali.

Dopo aver compilato la maschera con i dati richiesti, cliccando su *next* si apre una pagina con due opzioni (Figura 4.3): la ricerca dei testi direttamente in Internet per mezzo del sistema informatico (opzione 1) o il caricamento nel sistema dei dati primari collezionati in precedenza (opzione 2). Una volta creato e/o caricato il corpus, si possono utilizzare alcune delle funzionalità presentate nel capitolo 2.

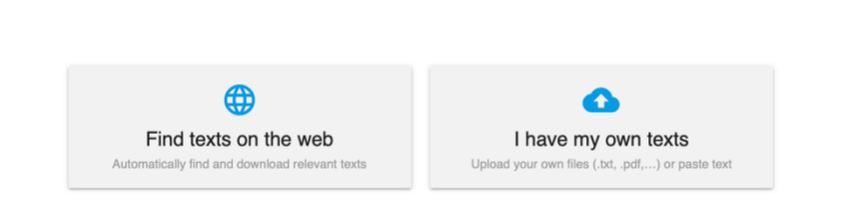


Figura 4.3: Screenshot delle opzioni presenti in *create corpus* in *Sketch Engine*

#### 4.2.1 Opzione 1

Se non si dispone dei dati primari per costruire il corpus, o se si vogliono integrare i dati utilizzando il materiale reperito in Internet, si può usare l'opzione *Find texts on the web*, che permette di trovare automaticamente i testi in rete e di scaricarli direttamente nel sistema. Cliccando sul riquadro di sinistra (Figura 4.4) si apre una schermata in cui si sceglie tra:

- la ricerca in rete (*Web search*): in questo caso, devono essere indicate da tre a 20 parole singole o multiple. Se l'obiettivo è la costruzione di un corpus legato a una particolare tematica, la scelta ricade su parole collegate alla tematica. Nel caso della migrazione, ad esempio, possono essere inserite parole come 'migrazione', 'rifugiato', 'profugo', 'migrante', 'crisi', 'accoglienza', 'Convenzione di Ginevra', 'centro di accoglienza' in lingua

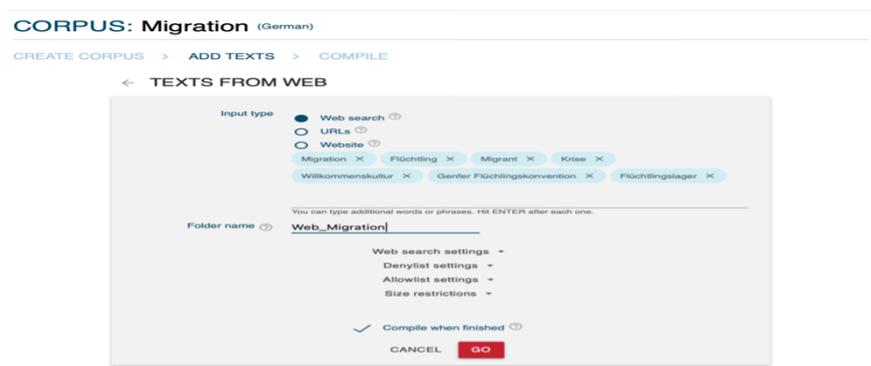
italiana oppure *Migration*, *Flüchtling*, *Migrant*, *Krise*, *Willkommenskultur*, *Genfer Flüchtlingskonvention*, *Flüchtlinglager* in lingua tedesca. Il sistema avvia una ricerca casuale delle parole date, procedendo in gruppi di tre, nel motore di ricerca *Bing*, e scarica in automatico le pagine web in cui compaiono queste parole;

- la ricerca per URL: si devono indicare gli indirizzi delle pagine web (un URL per linea). Per esempio, se si volesse costruire un corpus di guide turistiche, si potrebbero digitare gli URL di un numero di pagine web pertinenti, verificando però sempre che le pagine scaricate contengano materiale; dato che alcuni URL bloccano, tra gli altri, i *search robots*, le pagine scaricate potrebbero essere vuote;
- la ricerca per sito web (*Website*): con questa opzione si scarica tutto il sito web (con un limite di 10.000 pagine per sito). Indicando una *subdirectory*, si può però scaricarne anche solo una parte. Anche in questo caso sussiste il rischio di scaricare pagine vuote.

Una volta scelta una delle tre opzioni, occorre solo nominare la cartella in cui salvare i dati e si potrà poi procedere<sup>3</sup> cliccando su *go*.

Si illustra ora nel dettaglio la prima opzione (ricerca dei testi in rete), immaginando di voler costruire un corpus monolingue (lingua tedesca) sulla tematica della migrazione. Per esemplificare la procedura si utilizzano le seguenti parole: *Migration*, *Flüchtling*, *Migrant*, *Krise*, *Willkommenskultur*, *Genfer Flüchtlingskonvention*, *Flüchtlinglager*.

Una volta inserite le parole singole e multiple scelte, cliccando sul tasto *enter* per ognuna, il sistema (Figura 4.4) estrapola casualmente le loro combinazioni (Fig. 4.5):



**Figura 4.4:** Screenshot della pagina *texts from the web* dopo l'inserimento delle parole da cercare

3 È possibile specificare anche, in maniera opzionale, possibili parole da escludere (*Denylist settings*) o da ammettere (*Allowlist settings*), così come la grandezza del corpus (*Size restrictions*).



Figura 4.5: Screenshot delle triplette create da *Sketch Engine* dopo l'inserimento

Le combinazioni possono poi essere selezionate (lasciando la spunta prima della tripletta) o eliminate (eliminando la spunta prima della tripletta). Per ogni tripletta è possibile mantenere o eliminare le pagine web identificate (Figura 4.6):

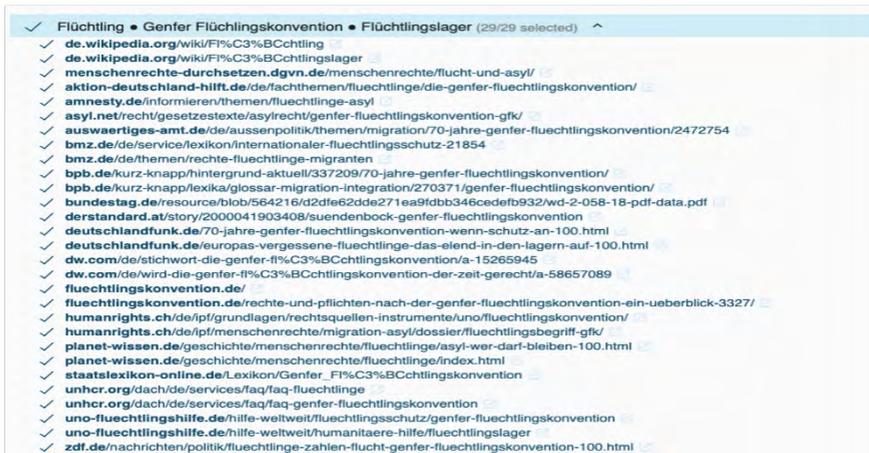


Figura 4.6: Screenshot delle pagine web estrapolate per la tripletta *Flüchtling – Genfer Flüchtlingskonvention - Flüchtlingslager*

Dopo il controllo delle pagine web e dopo aver scelto quelle adeguate agli obiettivi di ricerca (ad esempio, quelle esemplificative di un dato genere testuale), cliccando nuovamente su *go* viene dato avvio al processo di costruzione del corpus. La procedura richiede un certo tempo; nel frattempo, si può proseguire con il lavoro con il software usando altre funzionalità (caricando altri testi ecc.).

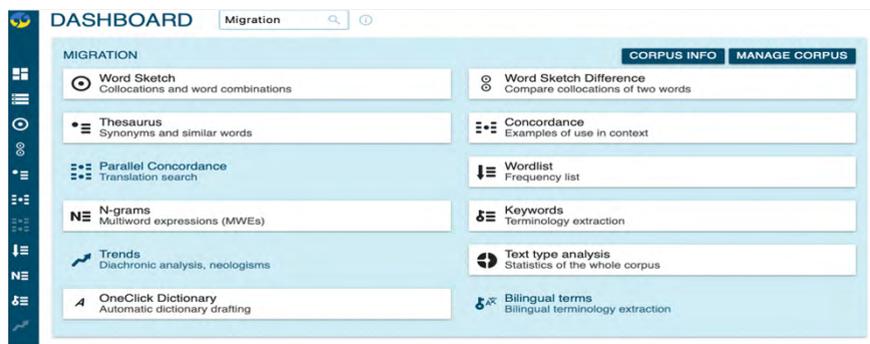
Una volta ultimato il caricamento dei dati nella cartella, il corpus può essere compilato, cliccando *compile*. Questa procedura avvia i processi automatici di tokenizzazione, lemmatizzazione e annotazione morfosintattica.

Il corpus creato<sup>4</sup>, denominato *Migration*, è ora pronto per essere utilizzato. Le caratteristiche principali sono consultabili sotto *corpus details and statistics* (una sintesi è presente nella Tabella 4.1).

<b>Token</b>	712.008
<b>Parole</b>	578.924

**Tabella 4.1:** Informazioni sul corpus *Migration*

Le varie funzionalità di analisi (*Word Sketch*, *Thesaurus*, *N-grams*, *OneClick Dictionary*, *Word Sketch Difference*, *Concordance*, *Word List*, *Keywords*, *Text Type Analysis* ecc.) sono visualizzate nel *Dashboard* (Figura 4.7) o nella barra laterale.



**Figura 4.7:** Screenshot del Dashboard di *Sketch Engine*

## 4.2.2 Opzione 2

Se è stata scelta l'opzione 2, che prevede il caricamento nel sistema di dati raccolti in precedenza<sup>5</sup>, allo step della Figura 4.3 si deve optare per *I have my own texts*.

Prima di illustrare in dettaglio questa procedura, è necessaria una premessa sui dati utilizzabili: vi sono, infatti, molte possibilità, come la realizzazione di documenti *ad hoc*, la scansione di documenti o il download da Internet (cfr.

4 I dati presenti nel corpus potranno poi essere incrementati in un secondo momento in base alle esigenze dell'analisi. Si potrà avviare una nuova ricerca con *Web search* o caricare i propri dati utilizzando l'opzione 2.

5 Chi raccoglie e carica i dati detiene anche il copyright e la responsabilità in *Sketch Engine*. I dati non potranno essere usati e divulgati dal sistema, a meno che la proprietaria/il proprietario del profilo decida di condividere i propri dati con altre persone o con la *community*.

Mukherjee 2002: 134). I dati, una volta salvati sul proprio dispositivo, devono essere raccolti in una cartella (in caso di un corpus monolingue) o in più cartelle (in caso di un corpus plurilingue)<sup>6</sup>. In ogni cartella i file devono essere nominati e categorizzati in maniera sistematica per poter essere identificati univocamente nel momento in cui andranno a confluire nel corpus. La descrizione dei dati primari (metadati) è una procedura molto importante: i metadati<sup>7</sup>, infatti, hanno la funzione di descrivere il design del corpus e determinano anche quali domande di ricerca potranno essere poste (cfr. Schmidt 2022a; Hunston 2008). Per esempio, per un corpus specialistico di testi turistici sarà di fondamentale importanza la tipologia testuale; per un corpus sulla migrazione in un certo arco temporale saranno centrali la tipologia testuale e la datazione; per un *learner corpus* avranno rilevanza la L1, la L2, il luogo, il tipo di prova ecc.

I dati possono poi essere caricati direttamente dal proprio computer cliccando sulla nuvoletta e scegliendo i file dal proprio strumento oppure trascinando i file direttamente nello spazio apposito<sup>8</sup>. Non possono essere caricati più di 100 file contemporaneamente e il peso dei dati non può superare i 500 MB. I formati ammessi sono vari: tra gli altri .csv, .doc, .docx, .htm, .html, .ods, .pdf, .txt, .zip.

Una volta caricati i dati si procede alla loro tokenizzazione, lemmatizzazione e annotazione morfosintattica automatica cliccando su *compile*. Le caratteristiche del corpus sono poi visibili sotto *corpus details and statistics*. Per iniziare l'analisi si apre il *corpus dashboard*, scegliendo una delle funzionalità presentate precedentemente (Figura 4.7).

---

6 Per la procedura di costruzione di un corpus comparabile in *Sketch Engine*, cfr. Flinz (2020: 143).

7 Esistono vari tipi di metadati (cfr. anche paragrafo 1.2). Schmidt (2022a: 249-251) distingue tra: *Korpusdesign-Metadaten*, che determinano il design di un corpus; *Katalog-Metadaten*, che descrivono le caratteristiche principali di un corpus; *organisatorische Metadaten*, necessari per la costruzione o il salvataggio di un corpus in un sistema informatico.

8 È possibile copiare anche il testo usando l'opzione *paste text*, anche se questa modalità è consigliata solo per testi brevi.



## PARTE II

DESCRIZIONE DI FENOMENI MORFOLOGICI E SINTATTICI  
BASATI SULL'USO DI CORPORA E APPLICAZIONI



# Capitolo 1

## Morfologia

In questo capitolo il focus d'indagine è sulla morfologia, ossia su quella branca della linguistica che si occupa della «Gestalt, Flexion (Beugung) und Bildung von Wörtern» (Busch/Stenschke 2018: 80). Partendo da una sintetica introduzione all'oggetto della morfologia e dei suoi ambiti di ricerca, si mostra per mezzo di due casi studio come i corpora possano essere utilizzati per illustrare determinati fenomeni, per formulare ipotesi e operare sia in maniera induttiva che deduttiva. Seguono alcune attività applicative.

### 1.1 Introduzione

#### 1.1.1 La parola e i suoi elementi costitutivi

Oggetto di studio della morfologia è la struttura interna della parola (*das Wort*)<sup>1</sup> (dal greco *morphḗ*). Definire cosa è una parola non è un compito semplice, dal momento che i suoi significati sono vari. Anche se le definizioni in letteratura sono molteplici<sup>2</sup> (cfr. Busch/Stenschke 2018: 80), possono essere stabiliti almeno cinque criteri fondamentali:

1. il criterio ortografico (*ortographisches Kriterium*), per il quale la parola è l'unità linguistica compresa tra due spazi (cfr. Busch/Stenschke 2018: 81);
2. il criterio fonetico-fonologico (*phonetisch-phonologisches Kriterium*) secondo cui la parola è caratterizzata da un'identità prosodica. Le parole, infatti, sono dotate di particolari caratteristiche di accento<sup>3</sup>, come si può evincere dalla differenza tra le due sequenze omofone *Meerwasser* (“acqua di mare”)<sup>4</sup> e *mehr Wasser* (“più acqua”)<sup>5</sup> (cfr. Di Meola 2014: 58);
3. il criterio lessicale-semantico (*lexikalisch-semantisches Kriterium*), per il quale la parola è dotata di un'autonomia semantico-funzionale, ossia ha un significato o svolge una funzione grammaticale (cfr. Busch/Stenschke 2018: 81);
4. il criterio morfologico (*morphologisches Kriterium*), secondo il quale la parola preserva in tutte le forme la sua struttura fondamentale, che rimane stabile e non può essere interrotta da altri elementi;
5. la mobilità sintattica (criterio sintattico, *syntaktisches Kriterium*), tale per cui la parola è l'elemento più piccolo che può essere spostato in una frase.

1 Il plurale di *Wort* è duplice: *Worte* e *Wörter*. Cfr. <https://www.dwds.de/wb/Wort> [10.2.23].

2 Si vedano anche le definizioni offerte dai dizionari di linguistica tedesca (cfr. Bußmann 2008; Glück/Rödel 2016).

3 Busch/Stenschke (2018: 81) mettono in discussione questo criterio, dato che esistono delle unità come *der*, *mir*, *die* che non hanno accento e che non potrebbero essere quindi considerate parole.

4 Le traduzioni in lingua italiana sono ad opera dell'autrice, salvo diversa indicazione.

5 Cfr. *Meerwasser* ['me:ɐ̯vasɐ] e *mehr Wasser* ['me:ɐ̯'vasɐ].

Secondo la *Dudengrammatik* le parole sono «elementare [ ] Einheiten, aus denen nach den Regeln der Syntax komplexere sprachliche Einheiten, nämlich Phrasen und schließlich Sätze, aufgebaut werden» (Duden 2022: 595).

Dalla parola, in un certo senso concreta, che può assumere forme diverse (*Wortformen*), come per esempio *gehen*, *geht*, *ging*, *gegangen*, si distingue il lessema (*Lexem*), chiamato anche parola lessicale (*lexikalisches Wort*), ossia l'unità astratta che fa parte del lessico di una lingua. La relazione tra forma astratta e concreta è assimilabile alla relazione type (forma astratta) – token (forma concreta) (cfr. Lemnitzer/Zinsmeister 2015: 198). Ogni type può infatti assumere, nell'uso concreto, una forma o una funzione sintattica diversa (cfr. Lemnitzer/Zinsmeister 2015: 198). Il lessema, inoltre, è quello che viene inserito nel dizionario per mezzo della sua forma base, la *Nennform* (cfr. Duden 2022: 596); in lessicografia si usa il termine “lemma” (*Lemma*), o anche “voce” o “entrata lessicale” (*Stichwort*, *Eintrag*): i verbi sono in genere riportati all'infinito (*singen* per la forma *singt*), i nomi e i pronomi alla forma del nominativo singolare (*Hund* per la forma *Hunde*) e l'aggettivo alla forma utilizzata per esprimere il predicativo<sup>6</sup> (*die Schube sind neu*). Il livello astratto e il livello concreto di una parola sono ben visibili in alcune preposizioni del tedesco, come per esempio in *zur*, che ha la forma di una parola unica, ma, di fatto, è costituita da due lessemi distinti: *zu* + *der*.

Le parole vengono normalmente suddivise in classi di parola (*Wortarten*), dette anche classi grammaticali o parti del discorso. Le classi di parola si determinano in base a vari criteri, come:

1. il criterio morfologico, che distingue le parole in classi variabili e invariabili. Nelle classi variabili sono comprese le parole coniugabili come i verbi, che possono essere modificati in base alla persona, il numero, il tempo e il modo, e le parole declinabili come il nome, l'aggettivo, l'articolo, il pronome, che possono variare in base al genere, numero e caso. Le classi invariabili comprendono parole che non si modificano, come l'avverbio, la preposizione, la congiunzione e la particella (cfr. Tabella 1.1):

Classe	Tipo di parola	Esempi
parole variabili	parole coniugabili	verbo
	parole declinabili	nome, aggettivo, articolo, pronome
parole invariabili	parole che non si modificano	avverbio, preposizione, congiunzione, aggiuntore, particella, interiezione

**Tabella 1.1:** Classi di parola in base al criterio morfologico<sup>7</sup>

<sup>6</sup> Per gli aggettivi che possono essere usati solo in forma attributiva, si utilizza solitamente la forma debole in *-e* (cfr. Duden 2022: 597). È il caso di *vordere*, che finisce in *-e* in *der vordere Eingang*.

<sup>7</sup> Si veda anche il paragrafo 2.1.1 (Parte II).

2. il criterio numerico, che differenzia tra classi aperte (verbo, nome, aggettivo, avverbio) e classi chiuse (articolo, pronome, preposizione, congiunzione, particella). Le prime sono classi ampie ed espandibili grazie all'apporto di neologismi e occasionalismi; le seconde racchiudono invece un numero di parole costante nel tempo;
3. il criterio semantico, che distingue tra classi lessicali, che comprendono parole con un significato lessicale pieno (nomi, verbi, aggettivi, avverbi), e classi funzionali, in cui le parole hanno un significato primariamente grammatico-relazionale (articoli, pronomi, preposizioni, congiunzioni e particelle).

Una parola può essere semplice da un punto di vista morfologico (*Simplex*), quindi non più suddivisibile in unità dotate di significato, oppure complessa (*Wortbildungskonstruktion/Wortbildungsprodukt*), dunque composta da unità diverse dotate di significato, o almeno portatrici di una caratteristica grammaticale (cfr. Busch/Stenschke 2018: 82). Queste unità elementari da un punto di vista morfologico, che hanno o un significato o una funzione grammaticale, vengono definite morfi (*Morph*)<sup>8</sup> e si indicano tra parentesi graffe: {Haus}, {e}.

Nella lingua tedesca una stessa sequenza di suoni (ossia lettere), come *Brücke*, può essere usata in combinazione con significati diversi in morfi diversi (cfr. Busch/Stenschke 2018: 83). È opportuno, a questo proposito, accennare al fenomeno della polisemia (*Polysemie*) e della omonimia (*Homonymie*):

- Per polisemia si intende una relazione semantica tra lessemi che hanno significati diversi, ma la stessa forma. I significati sono determinati da una radice etimologica comune come il caso di *Wurzel* che si usa per piante, denti, numeri, eccetera<sup>9</sup>.
- Per omonimia si intende una relazione semantica tra lessemi che hanno la stessa forma, ma significati diversi; tra i due non vi è però alcuna radice etimologica in comune come nel caso di *Schloss* (castello vs. serratura)<sup>10</sup>.

Al contrario, una serie di significati o caratteristiche lessicali può essere espressa con morfi diversi, ossia allomorfi (*Allomorph*)<sup>11</sup> (1):

- (1) Allomorfi del morfema per la 2. persona singolare: {st} come in *wohn-st*, {est} come in *arbeit-est*

In alcuni contesti è possibile un solo allomorfo, in altri due o più. La *Dudengrammatik* (cfr. Duden 2022: 607) riporta il caso degli allomorfi del morfema per il genitivo di nome maschile e neutro, {s} e {es}, dove per esempio per *Garten* ricorre solo {s} (cfr. *des Gartens*) mentre per *Turm* sia {s}, sia {es} (cfr. *des Turms* e *des Turmes*).

8 Al singolare morfo (*Morph*).

9 Cfr. <https://grammis.ids-mannheim.de/terminologie/534> [30.08.23].

10 Cfr. <https://grammis.ids-mannheim.de/terminologie/510> [30.08.23].

11 Cfr. anche la distinzione tra fonema e allofono in fonetica e fonologia.

L'unità costituita dal significato o dalla caratteristica grammaticale (il contenuto del segno) e il suo morfo/i suoi morfi, i suoi allomorfi (ossia l'espressione del segno), si definisce morfema (*Morphem*). Il morfema è l'unità minima di parola dotata di significato (*bedeutungstragende Einheit*)<sup>12</sup>: «Das Morphem ist die kleinste sprachliche Einheit mit einer bestimmten Form und einer bestimmten Bedeutung» (Duden 2022: 605). Anche il morfema si indica in parentesi graffa.

I morfemi possono essere di vario tipo:

- in base al significato si distingue tra morfemi lessicali (*lexikalische Morpheme*), che hanno un significato pieno, e morfemi grammaticali (*grammatische Morpheme*), portatori di informazioni grammaticali. Nella parola *Abendkleidung* vi sono tre morfemi: due lessicali {Abend}, {Kleid} e uno grammaticale {ung};
- in base alla loro distribuzione si differenzia tra morfemi liberi (*freie Morpheme*, chiamati anche *Grundmorpheme*, *Basismorpheme*, *Wurzeln*), che possono fungere da parola autonoma, e morfemi legati (*gebundene Morpheme*, *Wortbildungsmorpheme*), che possono comparire solo insieme a un morfema libero. I morfemi legati, a loro volta, possono essere classificati in base alla funzione e alla posizione nella parola (cfr. Pittner 2016: 60).

Queste tipologie di morfema spesso però si possono anche sovrapporre: un morfema lessicale è solitamente un morfema libero<sup>13</sup>, mentre i morfemi grammaticali sono sempre morfemi legati.

I morfemi si possono suddividere in radice (*Stamm*) e affisso (*Affix*). La radice è la parte principale della parola, ossia l'elemento che rimane una volta tolti tutti gli elementi non indispensabili: se si prende come esempio la parola *Frauen* (*Frau-en*), la radice è {Frau}, a cui segue il morfema per il plurale {en}. La radice, inoltre, è portatrice del significato della parola ed è formata di norma da un morfema libero; in parole complesse vi possono essere anche due o più radici (cfr. *Atomkraftwerke* = *Atom-kraft-werk*, cfr. Duden 2022: 603). Radice e base, ossia la forma da cui prendono avvio i processi di formazione delle parole, possono coincidere, ma la base può essere anche più ampia della radice. Di Meola (2014: 62) esemplifica questo processo attraverso la parola *Ent-fremd-ung*, la cui radice è *fremd*, mentre la base è *entfremd(en)*:

(2) dall'aggettivo *fremd* → al verbo *entfremd(en)* → al sostantivo *Entfremdung*

Gli affissi sono invece morfemi grammaticali utilizzati per formare altre

12 Nella distinzione fono/fonema il fonema è l'unità più piccola distintiva di significato (*bedeutungsunterscheidende Einheit*) (cfr. Busch/Stenschke 2018: 83).

13 Fanno eccezione i cosiddetti *unikale Morpheme* ("morfemi unici", chiamati anche "morfemi cranberry") che non hanno più significato e occorrono solo in un'unica parola come componente (come, per esempio, l'elemento {Brom} in *Brombeere*) e altri residui di morfemi liberi, come {Schwieger} in *Schwiegervater* e *Schwiegermutter*.

parole. In base alla loro posizione rispetto alla base si classificano in:

- prefissi, che precedono la base. Esempi: {un} in *un-sicher*, {ver} in *ver-lau-fen*, {un} in *Un-lust*;
- suffissi, che seguono la base. Esempi: {heit} in *Frech-heit*, {lich} in *lieb-lich*, {st} in *geb-st*;
- circonfissi<sup>14</sup>, che precedono e seguono contemporaneamente la base. Esempi: {ge + t} nei participi passati come in *ge-kauf-t*.

Nella categoria degli affissi rientrano anche i confissi<sup>15</sup> (*Konfixe*), che hanno un significato più concreto e sono costituiti da forestierismi, come {thek} in *Video-thek*. I confissi non compaiono solo in parole straniere, come per esempio in *Biodynamik*, ma anche in parole che combinano elementi autoctoni e stranieri (*Bioladen*) (cfr. Busch/Stenschke 2018: 96). Possono unirsi anche tra di loro (*biomorph*). Sempre negli affissi rientra anche la categoria degli infissi (*Infixe*), come {un} in *funktions-un-tüchtig*<sup>16</sup>.

Gli affissi, in quanto morfemi grammaticali, hanno normalmente un significato più astratto. Per esempio, {un} e {in} hanno un significato di negazione; {anti} di opposizione; {vor} di anteriorità temporale. Uno stesso suffisso può anche avere significati diversi: {er} può indicare colui che compie un'azione (*Lehrer* da *lehren*), ma anche uno strumento (*Lüfter* da *lüften*). Si vedano le parole *Brillen Träger* (colui che porta gli occhiali) e *Hosenträger* (oggetto che tiene su i pantaloni, ossia la bretella) (cfr. Di Meola 2014: 63).

Qui di seguito una sintesi dei tipi di morfemi sopradescritti (Tabella 1.2):

Tipi di morfema			
Libero		Legato	
Lessicale	Grammaticale	Lessicale	Grammaticale
verbale {geh}	pronominale {jemand}	verbale {helf}	circonfisso {ge+t}
nominale {Haus}	preposizionale {auf}	confisso {thek}	suffisso {t/et}
aggettivale {neu}	congiunzione {und}	prefisso {un}	prefisso {ver}
avverbiale {oft}		circonfisso {ge + t}	circonfisso {be + t}
		suffisso {ei}	suffisso {ung}

**Tabella 1.2:** Tipi di morfemi (riadattata da Busch/Stenschke 2018: 87)

14 In letteratura, i circonfissi vengono considerati come la combinazione di prefisso e suffisso (cfr. Di Meola 2014: 62).

15 Il loro status e la loro accettazione non sono sempre condivisi in letteratura (cfr. Busch/Stenschke 2018: 86). Gaeta (2017: 148) li suddivide in pre-confissi (*Präkonfixe*), che occorrono in prima posizione come {biblio} in *Bibliograph*, e in post-confissi, che compaiono come secondo elemento {skop} com *Bronchoskop*. Alcuni confissi possono occorrere in entrambe le posizioni {phil} come in *Philantrop* o *anglophil*.

16 *Funktions-un-tüchtig* potrebbe essere considerato anche come una parola composta: *Funktion + untüchtig*.

Dal momento che non è sempre facile attribuire un morfema a una classe, si preferisce non dare criteri rigidi, ma considerare fluidi i confini tra le varie classi (cfr. Busch/Stenschke 2018: 86).

### 1.1.2 Costruzione delle parole

Le parole possono essere costruite per mezzo di due procedimenti diversi: la flessione (*Flexion*) e la formazione delle parole (*Wortbildung*). Il tedesco è una lingua flessiva (cfr. Pittner 2016: 61). In base a come i morfemi vengono combinati per formare le parole, si distingue tra lingue isolanti come il cinese, in cui a ogni parola corrisponde un morfema, lingue agglutinanti come il turco, in cui i morfemi si combinano in parole plurisillabiche, e lingue flessive come il latino, in cui i morfemi si combinano in parole.

#### 1.1.2.1 La flessione

La flessione si distingue in declinazione per i nomi, gli aggettivi, i pronomi e gli articoli, e in coniugazione per i verbi. È un procedimento generalizzato e regolare. La regolarità è evidente nei neologismi e negli occasionalismi: se un verbo entra come prestito da un'altra lingua, può essere coniugato facilmente come ogni altro verbo tedesco.

Nella flessione nominale sono determinanti tre categorie grammaticali:

1. il genere (*Genus*), che può essere maschile (*Maskulinum*), femminile (*Femininum*) e neutro (*Neutrum*). Di Meola (2014: 64-66) mette in evidenza come alcuni processi morfologici, fonetico-fonologici e semantici possano aiutare a stabilire il genere di una parola, anche se non esiste un sistema di regole per determinare il genere di tutte le parole (cfr. Duden 2022: 700). Dalla struttura morfologica si possono ricavare, tra le altre, le seguenti informazioni:

- i suffissi autoctoni {er}, {ling} e i suffissi stranieri {ant}, {iker}, {ismus}, {ist} danno origine a parole maschili;
- i suffissi autoctoni {ei}, {heit}, {keit}, {schaft}, {ung} e i suffissi stranieri {ität}, {ade} danno origine a parole femminili;
- i suffissi autoctoni {chen}, {lein}, {tum} e i suffissi stranieri {ing}, {ment} formano parole neutre.

Dalla struttura fonetico-fonologica possono essere ricavate altre informazioni, tra cui a titolo esemplificativo:

- i monosillabi con «[j] + consonante», con *dr-*, *tr-*, *kn-*, con «nasale + consonante» all'inizio di parola sono quasi tutti maschili;
- i monosillabi che finiscono in fricativa [f], [ç], [x], + occlusiva dentale sorda – *t* o in *-ur* sono quasi tutti femminili;
- i monosillabi che finiscono in *-elt* sono quasi tutti neutri.

Si evincono quindi delle forti regolarità, che possono essere d'aiuto nella memorizzazione dei generi. Nelle parole plurisillabiche sono femminili il

90% delle parole che finiscono in *-e* (cfr. *schwa* come in *Tante*), mentre sono maschili l'80% di quelle che terminano in *-en* (*Regen*). Di Meola (2014: 65) generalizza il tutto mettendo in luce che le parole monosillabiche e quelle che terminano in consonante tendono a essere maschili, mentre le parole che terminano in vocale tendono a essere femminili.

La semantica fornisce poi altre indicazioni: se il sostantivo si riferisce a un essere umano, è generalmente determinante il genere naturale<sup>17</sup> del referente. Un ulteriore aiuto è fornito da alcune tendenze; tutte le denominazioni di bevande alcoliche (eccetto la birra, *das Bier*) e di marche automobilistiche sono maschili; tutte le denominazioni di marche di motociclette sono femminili; tutte le denominazioni di marche di birra e di hotel sono neutre (cfr. Duden 2022: 703).

In alcuni casi i tre fattori (morfologico, fonetico-fonologico e semantico) possono però entrare in conflitto. In tal caso, il processo morfologico prevarrà su quello semantico e il fattore semantico su quello fonetico-fonologico. Inoltre, vi sono alcune parole con un genere oscillante, soprattutto i prestiti: *Korpus* (*der/das*), *Pub* (*der/das*) ecc.

2. il numero (*Numerus*), singolare (*Singular*) e plurale (*Plural*). Per formare il plurale esistono vari morfemi, come {e}, {en}, {n}, {er} e {s}; in alcuni casi, non vi è alcun morfema che segnali il plurale (morfema zero, *Nullallomorph*). La scelta del morfema<sup>18</sup> è spesso determinata dalla struttura morfologica e dal genere della forma singolare (cfr. Di Meola 2014: 67), il che consente di trarre alcune generalizzazioni<sup>19</sup>: i sostantivi terminanti in *-ei*, *-in*, *-heit*, *-keit*, *-schaft*, *-ung* formano il plurale inserendo {e} o {en} (in alcuni casi si assiste a un raddoppiamento della consonante finale, come in *Ärztin*, il cui plurale è *Ärztinnen*); i sostantivi terminanti in *-ig*, *-ich*, *-ling*, *-nis*, *-sal* formano il plurale inserendo {e} (anche in questo caso, si assiste talvolta a un raddoppiamento della consonante finale, come in *Kenntnis*, il cui plurale è *Kenntnisse*); i sostantivi terminanti in *-tum* formano il plurale in {er}; i sostantivi terminanti in *-ler*, *-en*, *-chen* creano il plurale senza suffisso. Alcune forme plurali comportano anche una modifica della vocale radicale, come per *Wald*, il cui plurale è *Wälder*. Vi sono inoltre sostantivi che hanno solo il singolare (*Obs*) e sostantivi che hanno solo il plurale (*Eltern*);
3. il caso (*Kasus*): il tedesco ha quattro casi, il nominativo (*Nominativ*), l'accusativo (*Akkusativ*), il dativo (*Dativ*) e il genitivo (*Genitiv*).

I sostantivi presentano un genere stabile e si possono flettere solo in base al caso e al numero; la radice rimane identica eccetto lievi variazioni, come

17 Per una classificazione più dettagliata si veda la *Dudengrammatik* (cfr. Duden 2022: 701).

18 Per una panoramica dettagliata si veda la *Dudengrammatik* (cfr. Duden 2022: 707).

19 Oltre a queste regolarità, Di Meola distingue anche ulteriori tendenze, ma meno comuni (2014: 67).

l'inserimento di una *Umlaut* (*Mann* è al plurale *Männer*) e modifiche di accento (*Dóktor* diventa al plurale *Doktóren*). Gli aggettivi, invece, si modificano anche in base al genere, seguendo il genere del sostantivo a cui si accompagnano. La declinazione dell'aggettivo si orienta all'articolo o al pronome che precede. Tradizionalmente si distinguono tre tipi di declinazione:

- debole (*schwach*) dopo l'articolo determinativo, come in *der schöne Anzug*, con le sole desinenze {e}, {en};
- forte, senza articolo, come in *schöner Anzug*, con le desinenze {er}, {es}, {em} oltre a {e} ed {en};
- mista, dopo l'articolo indeterminativo, come in *ein schöner Bleistift*, sempre con {en}, eccetto {er} per il nominativo singolare maschile e {es} per il nominativo e accusativo singolare neutro.

Nella flessione verbale (coniugazione, *Konjugation*) in tedesco si distinguono le seguenti categorie:

- persona (*Person*): 1.; 2.; 3.;
- numero (*Numerus*): singolare (*Singular*), plurale (*Plural*);
- tempo (*Tempus*): presente (*Präsens*), preterito (*Präteritum*), perfetto (*Perfekt*), piuccheperfetto (*Plusquamperfekt*), futuro (*Futur I*), futuro anteriore (*Futur II*);
- modo (*Modus*): indicativo (*Indikativ*), congiuntivo (*Konjunktiv*), imperativo (*Imperativ*);
- diatesi (*Genus verbi*): attivo (*Aktiv*), passivo (*Passiv*).

Le forme dei paradigmi verbali possono essere semplici e complesse. In quelle semplici, si aggiungono i suffissi alla radice, come per l'indicativo e il congiuntivo presente, per l'indicativo e il congiuntivo preterito, per l'imperativo, per l'infinito presente (*hassen*) e per il participio (*hassend, gehass*). Le forme complesse sono costituite da un verbo ausiliare e da una forma non finita.

Si distingue, inoltre, tra coniugazione debole (*schwach*), regolare, e coniugazione forte (*stark*), irregolare. Nei verbi deboli la radice rimane costante e il suffisso che si è aggiunge è {t} (Tabella 1.3):

Infinito	Preterito	Perfetto
sagen	sagte	er hat gesagt

**Tabella 1.3:** Esempio di paradigma dei verbi deboli

I verbi forti<sup>20</sup> sono caratterizzati da una modificazione della vocale radicale e il loro participio viene formato con il suffisso in {en} (Tabella 1.4):

<sup>20</sup> I verbi forti sono ca. 170 (cfr. Di Meola 2016: 69). Osservandoli da un punto di vista diacronico si è notata una diminuzione dei verbi forti, tesi supportata dal fatto che i nuovi verbi

Infinito	Preterito	Perfetto
essen	aß	er hat gegessen

**Tabella 1.4:** Esempio di paradigma dei verbi forti

Nella flessione nominale e verbale, oltre alla modifica “marcata” con affissi, possono verificarsi anche modifiche interne della radice; ne sono un esempio la metaforia (*Umlaut*), fenomeno comune a molti plurali, e l’apofonia (*Ablaut*), ossia la variazione sistematica delle vocali all’interno dello stesso paradigma, come nei verbi forti (*i-a-u* come in *finden – fand – gefunden*; *-ei -i - i* come in *reiten – ritt – geritten*).

Nella categoria della flessione è possibile far rientrare anche la comparazione degli aggettivi (cfr. Pittner 2016: 62), per cui si distingue tra forma base (*Grundform, Positiv*) – forma comparativa (*Vergleichsform, Komparativ*) – forma superlativa (*Superlativ*), come in *hell – heller – am hellsten*. Questa scelta non è sempre accettata in letteratura (la comparazione può essere, infatti, attribuita anche alla formazione di parola); tuttavia, vi sono indicazioni sia a favore, come la regolarità del fenomeno, sia contro, come il fatto che le forme comparative e superlative possano essere usate in unità complesse (*Höchsttemperatur*), cosa che non accade per le altre forme flesse (cfr. Pittner 2016: 62).

### 1.1.2.2 La formazione delle parole

La formazione delle parole (*Wortbildung*) viene definita come «Untersuchung und Beschreibung von Verfahren und Gesetzmäßigkeiten bei der Bildung neuer komplexer Wörter auf der Basis vorhandener sprachlichen Mittel» (Bußmann 2008: 795f). Gaeta specifica che si tratta, in realtà, di formazione di lessemi (*Lexembildung*), e che comprende sia l’analisi del lessico complesso esistente, sia la produzione di nuovi lessemi (2017: 129).

I principali procedimenti di formazione di parola in tedesco sono tre: la composizione (*Komposition/Zusammensetzung*) (1.), la derivazione (*Derivation/Ableitung*) (2.) e la conversione (*Konversion*) (3.):

#### 1. La composizione (*Komposition/Zusammensetzung*)

La composizione è il procedimento mediante il quale si uniscono due o più parole per formare una parola composta (*Kompositum*). È il meccanismo di formazione delle parole più produttivo in tedesco (cfr. Gaeta 2017: 131).

I composti possono essere occasionali (*Okkasionalismen, Ad-hoc-Bildungen, Gelegenheitsbildungen, Augenblicksbildungen*) o usuali. Nel primo caso si tratta di composti creativi, che non entrano nel lessico di una lingua, ossia non si lessicizzano. Nel secondo caso si tratta di composti entrati stabilmente nell’uso di

---

che entrano a far parte della lingua tedesca si coniugano secondo la coniugazione debole (cfr. Pittner 2016: 64).

una lingua e che pertanto non vengono più percepiti come nuovi (cfr. Duden 2022: 611).

Nel caso in cui il composto si fossilizzi, la parola diventa sempre meno trasparente; ciò accade per esempio quando un elemento del composto non esiste più come parola a sé stante o esiste solo come elemento unico. È il caso di {Him} in *Himbeere*. Questi morfemi vengono definiti “morfemi cranberry” (*unikale Morpheme*). Busch/Stenschke (2018: 95) distinguono anche tra morfemi attivi (*aktive Morpheme*), e morfemi produttivi (*produktive Morpheme*). Nel primo caso si tratta di quei morfemi che esistono ancora nella lingua contemporanea, ma che non si utilizzano più per la formazione delle parole. Un esempio è {ling}. Nel secondo caso sono morfemi che vengono usati per la formazione delle parole come per esempio {computer}.

I composti possono essere determinativi (*Determinativkomposita*) o copulativi (*Kopulativkomposita*) (cfr. Duden 2022: 630). Sono determinativi quando vi è una gerarchia tra gli elementi costitutivi: la testa (*Kopf*), ossia il determinato (*Determinatum*), è l'elemento principale (*Grundwort*), mentre il modificatore, il determinante (*Determinans*), è l'elemento secondario (*Bestimmungswort*). Il significato fondamentale viene espresso dalla testa, mentre l'elemento secondario ha la funzione di specificazione. Mentre in italiano l'ordine degli elementi usuale è «testa + modificatore» (si veda *capotreno*, dove *capo* è la testa e *treno* il modificatore)<sup>21</sup>, in tedesco l'ordine è «modificatore + testa», come in *Haus Schlüssel*, dove *Haus* è il modificatore e *Schlüssel* la testa. La testa non è solo portatrice di significato (*Kern des Kompositums*), ma anche delle caratteristiche grammaticali dell'intero composto. Si prenda ad esempio la parola composta *Weinflasche*, dove *Flasche* è la testa e *Wein* il modificatore. *Wein* specifica il contenuto della bottiglia (si vedano anche *Wasserflasche*, *Bierflasche* ecc.). Il genere della parola composta è dato, invece, da *Flasche* (*die*), come è sempre la parola *Flasche* a determinare il plurale del composto (*Weinflaschen*). La testa, quindi, determina la classe di parola e, per i sostantivi, anche il genere e la classe flessiva.

Qui di seguito un riepilogo della terminologia in tedesco (Tabella 1.5):

Haus Schlüssel	
Haus	Schlüssel
Erstglied	Zweitglied
Bestimmungswort	Grundwort
Determinans	Determinatum
Modifikator	Kopf/Kern

**Tabella 1.5:** Riepilogo dei termini usati in tedesco per identificare le unità in un composto determinativo (cfr. Pittner 2016: 67)

21 Cfr. anche Bosco Coletso/Costa (2013: 120). In alcuni casi vi sono delle eccezioni, come per *autostrada* (cfr. Di Meola 2014: 73) o quando sono parole composte non autoctone (cfr. Gredel/Flinz 2020; Flinz/Gredel 2022).

I composti determinativi (*Determinativkomposita*) possono essere anche molto complessi in tedesco; in alcuni casi possono essere anche intere frasi (*Leck-mich-am-Arsch-Haltung*). I composti complessi sono poi suddivisibili per unità vicine, in maniera binaria, raggruppando le basi lessicali a due a due: *Motorradweltermeisterschaft* può essere suddiviso in primis in *Motorrad* | *Weltmeisterschaft*. L'analisi in costituenti immediati (*IC-Analyse*, *Analyse der unmittelbaren Konstituenten*) permette di scomporre una parola composta nelle sue unità per mezzo di diagrammi ad albero (*Baumdiagramme*) o di parentesi (*Klammerung*) (cfr. Pittner 2016: 67; Busch/Stenschke 2018: 88-96)<sup>22</sup>.

I composti copulativi (*Kopulativkomposita*), invece, non sono caratterizzati a livello semantico da un rapporto gerarchico: *süß-sauer*, *rot-grün*. Le due unità rappresentano parti di un tutto e sono entrambe portatrici di significato. Come si evince dai casi sopra illustrati, l'ordine non è fisso e gli elementi sarebbero potenzialmente interscambiabili (*süß-sauer* / *sauer-süß*; *rot-grün* / *grün-rot*). In realtà, per alcune parole uno dei due ordini è diventato più comune, ossia più convenzionale. Anche l'accento dei composti copulativi si distingue da quello dei composti determinativi: nei primi può cadere su entrambe le unità o sul secondo elemento, mentre nei secondi necessariamente sul primo elemento (cfr. Pittner 2016: 68). Vi sono addirittura casi in cui una parola composta può essere sia un composto determinativo sia un composto copulativo. In tal caso, il significato è distinguibile per mezzo dell'accento:

(3) BLAU-weiß ('bläulich weiß') vs. blau-WEISS ('blau und weiß')

I composti copulativi sono molto più rari dei composti determinativi e hanno meno possibilità di combinazione: le unità, per esempio, devono appartenere alla medesima classe di parola e alla medesima sottoclasse semantica (colori, partito politico ecc.). Le unità possono essere anche tre o più: *schwarz-rot-gelb*.

Sia i composti determinativi che i composti copulativi sono di tipo endocentrico, perché il nucleo semantico coincide con la testa (cfr. Gaeta 2017: 132).

Un altro tipo di composto citato in alcuni volumi di riferimento è il composto possessivo (*Possessivkompositum*)<sup>23</sup>, come *Eierkopf* ('jemand, der einen Eierkopf hat'), detto anche composto esocentrico. Nei composti esocentrici il nucleo semantico è da ricercare in un referente esterno al composto; né il primo elemento, né il secondo elemento sono ricollegabili al significato (cfr. Pittner 2016: 69): *Dummkopf* e *Rothaut* non si riferiscono a una parte del corpo, ma a

22 Busch/Stenschke (2018: 88-96), dopo aver descritto la procedura di analisi, riflettono su alcune sue criticità. In particolare, si soffermano sulle problematiche relative all'analisi dei circonfissi e alla suddivisione delle parole in cui sono presenti morfemi poco attivi e produttivi o morfemi attivi ma poco produttivi. Il cambiamento linguistico ha portato, infatti, alla modifica della funzione di alcuni suffissi, come per esempio {ling} (cfr. Eisenberg 2013a).

23 Il composto possessivo è un composto determinativo (cfr. Pittner 2016: 69).

una persona con la caratteristica X. *Dummkopf* è quindi una persona stupida (cfr. *dummer Mensch*)<sup>24</sup> e *Rothaut* è la denominazione dell'indiano del Nordamerica<sup>25</sup>.

Anche la classe di parola degli elementi del composto può essere un criterio di categorizzazione: un composto è nominale (N) quando la testa è un nome; un composto è verbale (V) quando la testa è un verbo; un composto è aggettivale (A) quando la testa è un aggettivo; un composto è avverbiale (Avv) quando la testa è un avverbio. I più diffusi sono i composti nominali (N), meno diffusi sono quelli aggettivali (A) e rari quelli avverbiali (Avv). Le diverse tipologie di composti possono essere ordinate in base alla loro frequenza (dal più frequente al meno frequente) (cfr. Di Meola 2014: 75; Duden 2022: 631):

1. composti nominali: Nome + Nome (*Tennisschub*), Aggettivo + Nome (*Schnellimbiss*), Verbo + Nome (*Bratwurst*), Preposizione + Nome (*Gegenangriff*), Avverbio + Nome<sup>26</sup> (*Nichtraucher*), Pronome + Nome (*Ichform*), Numerale + Nome (*Zweikampf*), Interiezione + Nome (*Aha-Erlebnis*), Confisso + Nome (*Mikrodarleben*), Gruppo di Parole + Nome (*Trimm-dich-Pfad*);
2. composti verbali: Avverbio + Verbo (*entgegenkommen*); Preposizione + Verbo (*durchdrehen*), Aggettivo + Verbo (*freihalten*), Nome + Verbo (*danksagen*);
3. composti aggettivali: Nome + Aggettivo (*hilfsbereit*), Aggettivo + Aggettivo (*hellgrün*), Verbo + Aggettivo (*denkfaul*), Preposizione + Aggettivo (*übernervös*), Altri (Avverbio + Aggettivo come *linksradikal*, Pronome + Aggettivo come *ichbezogen*, numerale + Aggettivo come *viereckig*, Confisso + Aggettivo come *multinational*, Gruppo di Parole + Aggettivo come *Lilien- und Astern-blau*);
4. composti avverbiali: Avverbio + Avverbio (*überallbin*), Aggettivo + Avverbio (*weither*), Preposizione + Avverbio (*übermorgen*).

Come accennato sopra, il composto più frequente è Nome + Nome. Le relazioni tra gli elementi possono essere svariate: il modificatore può indicare, infatti, il materiale, il luogo o la provenienza, la causa, lo strumento ecc. In diversi composti NN la testa è un sostantivo derivato da un verbo (transitivo o intransitivo): *Mathelehrer* (*Lehrer* da *lehren*), *Schneefall* (*Fall* da *fallen*). Il significato del composto si può ricavare direttamente:

- (4) *Mathelehrer* → jemand, der Mathe lehrt

24 “Dummkopf”, *Digitales Wörterbuch der deutschen Sprache*, <https://www.dwds.de/wb/Dummkopf> [15.06.2023].

25 “Rothaut”, *Digitales Wörterbuch der deutschen Sprache*, <https://www.dwds.de/wb/Rothaut> [15.06.2023].

26 La categoria AvvN (cfr. Di Meola 2016: 75) viene chiamata *Partikel + Nomen* in Pittner (2016: 66).

In questi composti, chiamati composti di reggenza o argomentali (*Rektionskomposita*), il rapporto semantico tra la testa (qui detta *Regens*) e il modificatore (*Rektum*) è lo stesso che sussisterebbe tra il verbo (*lehren*) e il complemento (*Mathe*) retto dal verbo nella corrispondente frase completa. Sono composti caratterizzati dalla presenza di una testa nominale deverbale (cfr. Gaeta 2017: 140). Per questo tipo di composto è possibile solo una interpretazione, a differenza di *Holzkeiste*, composto ambiguo<sup>27</sup> interpretabile sia come una scatola di legno, sia come una scatola per la legna. In alcuni casi, invece, la relazione tra gli elementi del composto NN è indiretta e non deducibile dalla semantica del verbo, come in *Geisterfahrer* (“l’automobilista che guida contro mano”).

Nella maggioranza dei composti nominali gli elementi sono uniti direttamente (*Nullfuge*), senza elemento di raccordo o interfisso (*Fugenelement*)<sup>28</sup>, ma questo dipende dal tipo di elemento coinvolto. È possibile, infatti, ricordare le unità di una parola composta con i seguenti morfemi: {s} (*Anfahrtsweg*), {n} (*Blumenstrauß*), {en} (*Schuldenschnitt*), {e} (*Hundehütte*), {er} (*Bilderrahmen*), {es} (*Kindesraub*), {ens} (*Schmerzengeld*), {ø} (*Buchseite*). La scelta di un morfema di raccordo è determinata dalle caratteristiche morfologiche, fonologiche e semantiche del modificatore (cfr. Duden 2022: 622); per esempio, in alcuni sostantivi coincide con la desinenza della forma plurale (*Kind-er-wagen*) o del genitivo singolare<sup>29</sup> (*Herz-ens-leid*). Le regole fisse per l’uso dei morfemi sono però esigue<sup>30</sup>; solo in alcuni casi, infatti, l’uso di un elemento di raccordo può essere collegato a un determinato suffisso, come p. es. *Versicherung-s-vertreter* (cfr. Pittner 2016: 71). Alcuni modificatori permettono vari elementi di raccordo (*Kindbett*, *Kindstod*, *Kindesalter*, *Kinderarzt*); in alcuni casi sono attestate per uno stesso lessema diverse possibilità, come in *Aschewolke*, *Aschenwolke*. Inoltre, sono presenti anche varianti diatopiche (*Schweinebraten*, *Schweinsbraten*).

Solitamente nel tedesco odierno i composti si scrivono uniti, anche se è possibile inserire un trattino (*Bindestrich*) come in (5) per distinguere i singoli elementi di un composto (cfr. Pittner 2016: 71).

27 Indipendentemente dalla loro tipologia, tutti i composti sono caratterizzati da indeterminata vaghezza semantica. Proprio questa caratteristica rende la composizione uno strumento espressivo molto forte (cfr. Gaeta 2017: 134).

28 L’origine dei morfemi di raccordo può essere fatta risalire a morfemi flessivi (cfr. Gaeta 2017: 134-135), anche se in letteratura vengono avanzati pareri sia favorevoli, sia contrari a questa interpretazione (cfr. Pittner 2016: 70-71). Nella lingua odierna i morfemi di raccordo sono però usati indipendentemente dalla loro origine flessiva (cfr. Di Meola 2014: 79).

29 In questo caso si parla di elementi di raccordo paradigmatici (*paradigmatische Fugen*). Nel caso si verificasse questa coincidenza, gli elementi di raccordo vengono definiti non paradigmatici (*unparadigmatische Fugen*) (cfr. Pittner 2016: 70): *Zeitung-s-anzeige*.

30 L’uso di un elemento di raccordo può essere collegato a un determinato suffisso: *Versicherung-s-vertreter* (cfr. Pittner 2016: 71).

## (5) Dritte-Welt-Herrscher

Il trattino è indice del processo di lessicalizzazione (cfr. Duden 2022: 612).

Nella pubblicità e in rete si sono attestate anche le seguenti modalità di scrittura: l'uso della maiuscola nel secondo elemento (*BahnCard*), oppure la separazione dei due elementi (*Kunden Service*).

L'accento della parola composta, che è il risultato dell'unione di almeno due lessemi con accento autonomo, è caratterizzato da una gerarchia prosodica tale per cui l'accento principale cade sul modificatore (principio del modificatore, cfr. Gaeta 2017: 137). Ampliando il composto, il principio del modificatore viene sostituito dal principio della testa complessa: l'accento si colloca sulla testa ramificata.

2. La derivazione (*Derivation/ Ableitung*)

Il processo di derivazione per la formazione di nuove parole si attua per mezzo di affissi<sup>31</sup> (suffissi, prefissi e circonfissi), morfemi legati, in combinazione con basi lessicali.

In base al tipo di affisso si distinguono i tre tipi di derivazione:

- derivazione con prefisso (*Präfigierung*): *ent-setzt*;
- derivazione con suffisso (*Suffigierung*): *Versicher-ung*;
- derivazione con circonfisso (*Zirkumfigierung*): *Ge-kreisch-e*.

Nella derivazione i suffissi e i prefissi hanno funzioni diverse:

- i suffissi determinano, per esempio, la classe di appartenenza della parola derivata, mentre i prefissi no;
- i suffissi autoctoni non sono mai portatori di accento principale<sup>32</sup>, i prefissi invece possono esserlo (cfr. Duden 2022: 619).

I principali suffissi<sup>33</sup> utilizzati per il processo di derivazione hanno soprattutto un significato grammaticale (Duden 2022: 619). La componente di significato<sup>34</sup> emerge solo in alcune coppie minime (*kindlich – kindisch*); alcune classi presentano un significato emotivo-affettivo, come i suffissi a carattere diminutivo {lein}, {chen} (cfr. Di Meola 2014: 81).

31 Gli affissi sono associati a un numero esiguo di significati derivazionali, raggruppabili in tre categorie: la trasposizione (passaggio di classe grammaticale della base); la modificazione (creazione di una denotazione caratterizzata da tratti semantici che modificano i tratti della base senza cambiarli, come per es. nella diminuzione); la mutazione (creazione di una denotazione caratterizzata da tratti semantici che cambiano i tratti della base) (cfr. Gaeta 2017: 145).

32 Eccezione è il prefisso {un}.

33 Se i suffissi contengono una -i- (come in -ig, -isch, -lich) possono produrre l'*Umlaut* nella base: {natur} + {lich} → *natürlich* (cfr. Duden 2022: 783).

34 Si vedano anche i suffissi elencati nella *Dudengrammatik* (cfr. Duden 2022: 724).

I suffissi determinano la classe di parola della parola derivata; si distinguono in: suffissi nominali (*Nominalsuffixe*), come in *Lehr-er*, *Heiz-ung*, *Gelegen-heit*, *Dumm-chen*, *Jüng-ling*; suffissi verbali (*Verbalsuffixe*), come in *läch-el-*, *amt-ier-*, *normal-isier-*, *stein-ig*, *krise-l-*; suffissi aggettivali (*Adjektivsuffixe*), come in *dümm-lich*, *rost-ig*, *herr-isch*; suffissi avverbiali (*Adverbsuffixe*), come in *nacht-s*, *haufen-weise*, *dumm-erweise*, *rück-lings*, *neuer-dings* (cfr. Pittner 2016: 72; Duden 2022: 618). Non tutti i suffissi si possono combinare con tutte le classi di parola; per esempio {in} può essere usato solo con sostantivi (*Freundin*). Alcuni suffissi possono essere combinati con varie classi di parola, come {er}, che viene unito a verbi (*erbeteln*) e ad aggettivi (*erblinden*) (cfr. Duden 2022: 618).

Anche nella derivazione, come nella composizione, domina il principio della testa a destra (*Kopf-rechts-Prinzip*) (cfr. Duden 2022: 617). I suffissi determinano non solo la classe di parola, ma anche il genere e la classe flessiva dei sostantivi (cfr. Pittner 2016: 73).

Alcuni esempi (cfr. Pittner 2016: 73-75):

- il suffisso nominale {er}: da verbi vengono derivati sostantivi maschili. Il sostantivo derivato identifica spesso una persona che svolge l'azione indicata dal verbo (*Nomen agentis*) o lo strumento utilizzato per svolgerla (*Nomen instrumenti*). In alcuni casi, anche l'azione stessa (*Nomen acti*): *Lehrer* (dal verbo *lehren*), *Bohrer* (dal verbo *bohren*), *Seufzer* (dal verbo *seufzen*). Il suffisso {er} è molto produttivo;
- il suffisso nominale {ung}: da verbi vengono derivati sostantivi femminili, che identificano il processo o il risultato dell'azione espressa dal verbo, come in *Verführung*;
- il suffisso nominale {in}: da sostantivi maschili, che identificano persone o animali, vengono derivati sostantivi femminili, come in *Lehrerin*;
- i suffissi nominali {heit}, {keit}, {igkeit}: da aggettivi vengono derivati sostantivi femminili. Esempi: *Schönheit*, *Dummheit*, *Wichtigkeit*;
- i suffissi nominali {ei}, {erei}: da nomi e verbi si derivano sostantivi femminili, come in *Bäckerei*. Si tratta di suffissi non autoctoni (di origine latina), dunque portatori di accento;
- i suffissi {chen} e {lein}: esprimono una riduzione della dimensione, anche in senso affettivo; questa può essere intesa sia in senso positivo, come in *Kätzchen*, sia negativo, con una punta di disprezzo, come nel caso di *Bürschchen*. Vengono usati per derivare sostantivi neutri: *Haus* – *Häuschen*. In rari casi si possono applicare anche ad altre classi di parola: *dumm* – *Dummchen*;
- il suffisso aggettivale {bar}: da un verbo, che di solito esprime un'azione, si derivano aggettivi, come per *essbar*. L'aggettivo esprime la fattibilità dell'azione: *essbar* (dal verbo *essen*) è un qualcosa di edibile. La derivazione da una base nominale, invece, non è più produttiva (*furchtbar*).

Qui di seguito la tabella riassuntiva dei suffissi (Tabella 1.6):

Suffisso	Base	Derivato
{er}	verbo, nome	nome
{ling}	verbo, nome, aggettivo	
{ung}	verbo	
{heit} {keit} {igkeit}	aggettivo, nome	
{e}	verbo, aggettivo	
{in}	nome	
{nis}	verbo, aggettivo	
{schaft}	nome	
{tum}	nome	
{ig}	nome, verbo, aggettivo, avverbio	
{lich}	nome, aggettivo, verbo	
{bar}	verbo, nome	
{haft}, {ig}, {isch}, {lich}, {los}	nome	
{ig}	nome	verbo
{ieren}	nome	
{s}, {weise}	nome	avverbio
{weise}, {erweise}	aggettivo	

**Tabella 1.6:** Selezione di suffissi di derivazione (da Di Meola 2014: 80; Pittner 2016: 75; Duden 2022: 618)

La derivazione per mezzo di prefissi non porta invece a una modifica della classe di parola (cfr. principio della testa a destra), ma solo del significato: *Unruhe*. In alcuni casi, i prefissi sono anche portatori di significato.

I prefissi sono generalmente atoni e neutri rispetto all'accento, che rimane sulla base lessicale, si vedano i prefissi verbali come {er}, {zer}, {ver}; i prefissi aggettivali e nominali, come per esempio {un}, {ur}, {miss}, sono invece accentati (cfr. Duden 2022: 619).

Qui di seguito un elenco di prefissi (Tabella 1.7):

Prefisso	Classe di parola	Tipo
{miss}, {un}, {ur}, {ent}	nome	
{un}, {ur}, {ent}	aggettivo	
{be}, {ent}, {er}, {ver}, {zer}	verbo	non separabile
{ab}, {an}, {auf}, {aus}, {ein}, {los}, {nach}, {vor}, {zu}		separabile
{über}, {durch}, {um}		separabile e non separabile <sup>35</sup>

**Tabella 1.7:** Alcuni prefissi di derivazione (da Di Meola 2014: 80; Pittner 2016: 75-76)

<sup>35</sup> In base al tipo di prefisso (separabile, non separabile), si evidenziano differenze di significato e di accento.

Il prefisso {un} può essere usato in aggettivi, in sostantivi così come in avverbi: *unschön* (aggettivo), *Unmensch* (sostantivo), *ungern* (avverbo). Il significato può modificarsi nel suo contrario (*Unlust*), può essere rafforzato, ampliato (*Unzahl*) o connotato negativamente (*Unwort*). Il prefisso {ent} si combina con basi verbali, aggettivali e nominali e indica, da un punto di vista semantico, una privazione, una mancanza, come nel caso di *entfremden*.

La derivazione per mezzo di circonfissi è più rara. Nei circonfissi il prefisso si combina con il suffisso, come per la modifica del sostantivo *Aufsicht* nel verbo *be-aufsicht-igen*. Possono essere derivati verbi da sostantivi, come in *be-schein-igen*, o da aggettivi, come in *be-schön-igen*; inoltre, si possono derivare sostantivi da verbi, come in *Ge-lauf-e*, e aggettivi da verbi, come in *un-erforsch-lich*. Il circonfisso più produttivo è *Ge-(e)*, usato per derivare basi nominali neutre da verbi. Mentre in passato era possibile assistere alla caduta della *e* (*Geschrei*), nelle formazioni più recenti la sua presenza è facoltativa, come per *Gepolter(e)*. Dal punto di vista del significato, queste ultime formazioni indicano un'azione ripetuta, vissuta negativamente.

Il confine tra composizione (unione di due o più morfemi liberi) e derivazione (unione di un morfema libero con morfemi legati) talvolta è poco chiaro. Gaeta (2017: 130) sottolinea che si tratta di un continuum, dove ai due poli si trovano i casi prototipici di composizione e derivazione e nel centro quelli meno prototipici, ossia i casi misti con proprietà intermedie. Per questi casi intermedi vengono spesso usati i termini affissoidi<sup>36</sup> (*Affixoide*, *Halbaffixe*), suffissoidi (*Suffixoide*) e prefissoidi (*Präfixoide*) in caso di desementizzazione. Si prenda ad esempio *Laubwerk*<sup>37</sup> (“fogliame”); nel tedesco moderno esiste il lessema *Werk*, ma il significato di {werk} “Produkt der Arbeit” non è più presente (cfr. Duden 2022: 623).

### 3. La conversione (*Konversion*)

La composizione e la derivazione sono caratterizzate dall'unione di due o più elementi; nella conversione invece non vi è alcun tipo di aggregazione, ma è la classe di parola che si modifica. La conversione è, infatti, il procedimento per il quale una base viene ricategorizzata, senza però l'ausilio di un suffisso<sup>38</sup>. I tipi di conversione (lessicale, morfologica) più diffusi in tedesco sono:

Verbo → Nome: la base verbale diventa un sostantivo, solitamente maschile. Questo procedimento è poco diffuso nella lingua contemporanea (cfr. Di Meola 2014: 84). Un esempio è: *beginn(en)* → *Beginn*;

Nome → Verbo: il sostantivo diventa base verbale. Questo procedimento è molto produttivo nella lingua contemporanea. Un esempio è: *Öl* → *ölen*;

36 Tra gli affissoidi vengono inseriti anche le cosiddette *Steigerungsbildungen*, come *saukalt* o *Affenhitze* (cfr. Pittner 2016: 78). Si veda anche Costa (2017).

37 Oltre a *-werk*, vengono citati *-reich*, *-arm*, *-riesen*, *-voll* (cfr. Duden 2022: 623).

38 Si utilizza anche l'espressione “derivazione mediante suffisso zero” (Di Meola 2014: 84).

Aggettivo → Verbo: l'aggettivo diventa base verbale. Un esempio è *reif*  
 → *reifen*;  
 Nome → Aggettivo: il sostantivo diventa aggettivo. Un esempio è *Feind*  
 → *feind*.

Pittner (2016: 80-81) distingue tra conversione lessicale e conversione sintattica. Mentre nella prima vi è il cambio di classe di parola della base, nella seconda è la forma di parola (base + flessivo) che cambia classe, come in *das Schwimmen*. Gaeta (2017: 160) parla in questo caso di nominalizzazione. Questo tipo di conversione viene chiamata anche trasposizione grammaticale (*grammatikalische Transposition*).

L'identificazione della direzione della conversione non è sempre trasparente (cfr. Duden 2022: 695), anche se segue alcuni criteri, come per esempio il tipo di prefisso, la forma, il significato ecc. (cfr. Pittner 2016: 79-80).

Altri procedimenti minori di formazione di parola, ma di natura più sporadica e meno sistematica sono (cfr. Duden 2022: 635):

- l'accorciamento o l'abbreviazione di parola (*Kurzwortbildung*);
- la contaminazione, ossia l'incrocio tra parole (*Kontamination/ Wortkreuzung/ Blendung/ Blending*);
- la formazione dei verbi separabili (*trennbare Verben, Partikelverben*);
- la reduplicazione/raddoppiamento (*Reduplikation*);
- la formazione di conglomerati (*Zusammenrückungen*).

L'accorciamento o l'abbreviazione di parola (*Kurzwortbildung*) è il procedimento mediante il quale la nuova parola si forma per mezzo di un accorciamento della base (*Ausdrucks-kürzung*)<sup>39</sup>. Si distingue normalmente tra *Kurzwörter*, ossia parole abbreviate sia dal punto di vista scritto che parlato e *Abkürzungen*<sup>40</sup>, parole abbreviate solo nello scritto (cfr. Leonardi 2012; Pittner 2016: 82-83; Busch/Stenschke 2018: 108). Nel parlato esse si leggono per esteso; esempi sono *usw.* (*und so weiter*) e *z. B.* (*zum Beispiel*).

Le abbreviazioni possono essere raggruppate sulla base di vari criteri. A seconda del tipo di abbreviazione si distingue tra:

- abbreviazioni che contengono solo un segmento della forma intera (*uni-segmentales Kurzwort*): *Zivi* (da *Zivildienstleistender*). Il segmento può provenire dall'inizio della parola, e in tal caso l'abbreviazione viene chiamata *Kopfwort* o *Initialwort* o *Anfangswort* (*Uni* da *Universität*). Questo tipo di abbreviazione è molto produttiva. In rari casi il segmento proviene dalla parte finale della parola, come *Rad* da *Fahrrad*. In questo caso

39 Tale procedimento si contrappone ai procedimenti presentati precedentemente, che rientrano invece nell'ampliamento di una parola (*Ausdrucks-erweiterung*) (cfr. Busch/Stenschke 2018: 108).

40 Per Pittner (2016: 82) e la *Dudengrammatik* (cfr. Duden 2022: 634) sono varianti di scrittura.

l'abbreviazione viene definita *Endwort* o *Schwanzwort*. In rarissimi casi ha origine dalla parte centrale della parola, come *Lisa* da *Elisabeth*; questo tipo di abbreviazione viene definita *Rumpfwort* (cfr. Busch/Stenschke 2018: 108).

- abbreviazioni che contengono più segmenti della forma intera (*multi-segmentales Kurzwort*): *Hivi* da *Hilfswissenschaftler*. In questo tipo di abbreviazione rientra l'acronimo (*Akronym*), costituito da una sequenza di lettere, come *EDV* da *elektronische Datenverarbeitung*. Si tratta di una tipologia di abbreviazione molto produttiva, soprattutto nell'informatica. Si pensi per esempio a *URL*, ormai molto diffusa, mentre la sua forma allungata *Uniform Resource Locator* è sconosciuta ai non esperti. Tali abbreviazioni possono essere pronunciate come sequenze di suoni (*FAZ* [fats] da *Frankfurter Allgemeine Zeitung*), come sequenze di lettere (*PKW* ['pe:ka:ve:] da *Personenkraftwagen*) o come forme miste (*WLAN* ['ve:la:n] da *Wireless Local Area Network*). In questo tipo di abbreviazioni rientrano anche l'abbreviazione in cui le sillabe iniziali stanno per i singoli elementi (*Silbenkurzwort*), come *Kripo* da *Kriminalpolizei*, e l'abbreviazione parziale (*partielles Kurzwort*) in cui si riduce solo la prima parte della parola, come *U-Bahn* da *Utergrundbahn*.

Le parole abbreviate mantengono solitamente il significato della parola intera. Talvolta hanno però un significato diverso, come nel caso di *Bafög* e *Bundesausbildungsförderungsgesetz*; infatti, si può dire che una studentessa/uno studente riceve il *Bafög*, ma non il *Bundesausbildungsförderungsgesetz*. Possono avere anche una connotazione diversa, come nel caso di *Nazi* – *Nazionalsozialist* o *Limo* – *Limonade*.

Nella contaminazione (*Kontamination/Wortkreuzung/Blendung/Blending*), considerata anche un particolare tipo di composizione, due parole si fondono insieme (cfr. Duden 2022: 635). Si possono unire il segmento iniziale e finale di due parole (una unità è però in forma abbreviata), come per esempio in *jein* (*ja* + *nein*) o *Denglisch* (*Deutsch* + *Englisch*). Spesso si tratta di neoformazioni occasionali con funzione scherzosa o satirica, come in *Maskne* (*Maske* + *Akne*), «durch Tragen einer Maske hervorgerufene Akne» (Duden 2022: 636). In rari casi la neoparola entra nell'uso, come per esempio *jein* o *Denglisch*; in altri è talmente convenzionalizzata da non essere più riconoscibile come conversione.

La categoria dei verbi separabili (*trennbare Verben, Partikelverben*), chiamata anche *Partikelverbbildung*, non è uniformemente accettata in letteratura; non si è concordi nel considerare i verbi separabili una parola. Gaeta inserisce questo fenomeno nella derivazione (2017: 142; 158-159), mentre la *Dudengrammatik* (cfr. Duden 2022: 635) lo considera un fenomeno a sé stante. Viene messa in luce la differenza tra *trennbare Verben/Partikelverben* e *Präfixverben*; i primi sono formati da particelle verbali, portatrici di accento, mentre i secondi da prefissi

verbalì (*Verbpräfixe*). I verbi separabili così seguono lo schema accentuale tipico dei composti, con accento sul modificatore, ossia la particella, mentre nei *Verbpräfixe* l'accento è sulla radice (cfr. Duden 2022: 692): 'anlächeln ("sorridere a qualcuno") – be'lächeln ("ridere di qualcuno/qualcosa"). Casi intermedi tra i prefissi verbali e le particelle verbali sono attestati mostrando un legame derivazionale tra i due (cfr. Gaeta 2017: 159).

La reduplicazione/raddoppiamento (*Reduplikation*) è un processo per il quale una parola o segmenti di parola vengono reduplicati (cfr. Duden 2022: 636), come per esempio *Blabla*. Frequente è la reduplicazione con variazione come in *Krimskrams* ("cianfrusaglie") o *ruckzuck* ("detto fatto").

La formazione di conglomerati (*Zusammenrückungen*) è caratterizzata dall'aggregazione di più parole legate a livello sintattico in una parola, come in *Vergissmichnicht* ("nontiscordardimé"), *Handvoll* ("manciata") (cfr. Fleischer/Barz 2012: 87; Foschi Albert/Hepp 2012: 123; Bosco Coletsos/Costa 2013: 123).

Altre modalità per formare nuovi lessemi sono il prestito (*Lehnwortbildung*) (cfr. Duden 2022: 940) e le espressioni multiparola (*Mehrwortlexeme*, *Mehrwortverbindungen*) (cfr. Duden 2022: 523).

Nel prestito si combinano elementi non nativi (affisso + radice, radice + radice) come in *transkulturell* o *Homeoffice*; esso si distingue dal calco strutturale (*Lehnübersetzung*), in cui il modello morfologico viene riprodotto usando materiale lessicale nativo in maniera più letterale (*Lehnübersetzung*), come in *Halbwelt* (da *demi-monde*), o in maniera più libera (*Lehnübertragung*), come in *Wolkenkratzer* (da *skyscraper*).

Le espressioni multiparola sono una modalità per arricchire il lessico di natura extramorfologica; possono essere più o meno fisse e vengono chiamate anche fraseologismi. Hanno un diverso grado di polilessicalità, fissità e idiomatilità (cfr. Burger 2015: 15-19). La polilessicalità può variare da un minimo di due lessemi fino ad arrivare a una frase intera. La fissità può andare dall'impossibilità di sostituire elementi alla presenza di variabili determinate dal contesto. Anche l'idiomatilità può essere totale o parziale.

Nella Tabella 1.8 è riportata una sintesi dei principali processi di formazione di parola in tedesco:

Formazione di parola ( <i>Wortbildung</i> )			
Tipo di modifica	Denominazione	Tipo di formazione	Esempi
ampliamento della base	composizione	composto determinativo	Hausschlüssel
		composto copulativo	süß-sauer
		composto nominale (N)	Tennisschuh (NN) Schnellimbiss (AN) Bratwurst (VN) Gegenangriff (PN) Nichtraucher (AVVN) Ichform (pronome + N) Zweikampf (numerale + N) Aha-Erlebnis (interiezione + N) Mikrodarlehen (confisso + N) Trimm-dich-Pfad (gruppo di parole + N)
		composto verbale (V)	entgegenkommen (AvvV) durchdrehen (PV) freihalten (AV) danksagen (NV)
		composto aggettivale (A)	hilfsbereit (NA) hellgrün (AA) denkfaul (VA) übernervös (PA) linksradikal (Avv + A) ichbezogen (pronome + A) viereckig (numerale + A) multinational (confisso + A) Lilien-und Astern-blau (gruppo di parole + A)
		composto avverbiale (Avv)	überallhin (AvvAvv) weither (AAVv) übermorgen (PAVv)
	derivazione	derivato (suffissi)	Versicher-ung
		derivato (prefissi)	ent-setzt
		derivato (circonfissi)	Ge-kreisch-e
	nessun ampliamento/ nessun accorciamento	conversione	V → N
N → V			Öl → ölen
A → V			reif → reifen
N → A			Feind → feind

accorciamento della base	abbreviazione unisegmentale	abbreviazione iniziale	Uni (da Universität)
		abbreviazione finale	Rad (da Fahrrad)
		abbreviazione centrale	Lisa (da Elisabeth)
	abbreviazione multisegmentale	acronimi	EDV (da elektronische Datenverarbeitung)
		abbreviazione (sillabe)	Kripo (da Kriminalpolizei)
		abbreviazione parziale	U-Bahn (da Untergrundbahn)
	contaminazione		Denglisch (Deutsch + Englisch)

**Tabella 1.8:** Principali processi di formazione di parola in tedesco

## 1.2 Morfologia con i corpora

In questa seconda parte si discutono due casi studio inerenti all'ambito linguistico della morfologia. Obiettivo è quello di esemplificare l'uso della linguistica dei corpora per l'analisi di aspetti morfologici; le fasi descritte possono essere prese a modello per l'analisi di ulteriori fenomeni.

Il primo caso studio ha come focus la composizione come mezzo per ampliare il lessico e per esprimere una valutazione (cfr. le funzioni della *Wortbildung* nella *Dudengrammatik* (cfr. Duden 2022: 609-610)); in particolare, si analizzano le parole composte che hanno *Klima* come modificatore in un corpus di CMC creato *ad hoc* sul discorso del cambiamento climatico (*Klimawandeldiskurs*). Obiettivo è l'identificazione delle parole composte più frequenti e gli *hapax legomena* per riflettere su occasionalismi, neologismi ed espressione di punti di vista.

Il secondo caso studio ha, invece, come oggetto di studio il confisso *-gate* e il suo status in un corpus di DeReKo<sup>41</sup>: il corpus *WP – Wikipedia Artikel*. Il confisso *-gate* è assai produttivo in molte lingue per formare composti per indicare uno scandalo. A livello diacronico nel tedesco scritto è emerso un cambiamento di significato.

### 1.2.1 Caso studio I: composti con *Klima* in un corpus di CMC costruito *ad hoc*

Una delle funzioni principali della *Wortbildung* è l'ampliamento del lessico (cfr. Duden 2022: 609), e la composizione è considerata uno dei suoi processi più produttivi (cfr. Glück/Rodel 2016; Konopka 2020). Altra funzione della *Wortbildung* è l'espressione di una valutazione o di un punto di vista (cfr. Duden 2022: 610).

41 Il corpus è *vdd19 – alle Wikipedia-Diskussionen zu Artikeln bis 2019*, facente parte dell'archivio *WP – Wikipedia Artikel u. Artikel- & Benutzerdiskussionen 2013/15/17/19*.

Nel presente studio entrambe le funzioni svolgono un ruolo centrale, dal momento che il focus d'indagine verte sul discorso del cambiamento climatico (*Klimawandeldiskurs*), oggetto di sempre più dibattiti, anche molto controversi. Gert Krell (2020) menziona i due poli estremi della discussione in atto, che oscilla tra allarmismo (*Alarmstimmung*), espresso per esempio dal composto *Klimahysterie* definito nel 2019 *Unwort des Jahres*, e la tendenza a sminuire il fenomeno (*Verharmlosung*), spesso accompagnata da atteggiamenti ottimistici (*Klimaschutzbemühungen*). Che la società nei vari paesi sia divisa è testimoniato anche da alcune indagini statistiche come, per esempio, quella svolta nel 2021 in Germania dall'organizzazione *More in Common*<sup>42</sup>.

Obiettivo del presente caso studio è quello di definire (1) quali siano i composti nominali più frequenti con il sostantivo *Klima* come modificatore e (2) quali siano gli *Hapax Legomena* (cfr. paragrafo 2.3) di un corpus CMC-costruito *ad hoc* sulla tematica scelta: il clima e il suo cambiamento.

Il corpus *ad hoc* (*Klima\_de*), contenente pagine web, post e commenti sui blog e social media, è stato creato con una delle funzionalità di *Sketch Engine* (*Find the text on the web*, si veda il paragrafo 4.2), utilizzando alcuni lessemi estrapolati da pagine specifiche in *Wikipedia* (cfr. Flinz/Luppi/Mell 2024): *Klima*, *Klimawandel*, *Erwärmung*, *Atmosphäre*, *Kohlenstoffdioxid*, *Treibhausgas*.

Il corpus, composto da 348.424 token, contiene dati tratti dai seguenti domini (Figura 1.1):

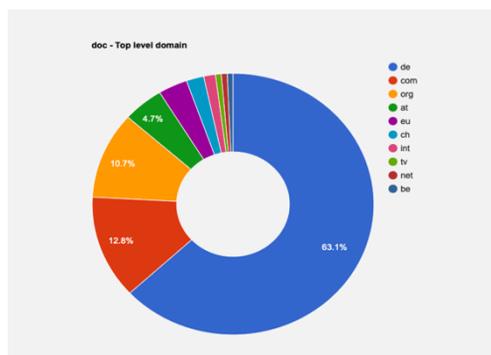


Figura 1.1: Domini web del corpus Klima\_de

L'approccio di analisi utilizzato è quello quantitativo-qualitativo (cfr. Lemnitzer/Zinsmeister 2015: 34): si utilizzano liste di frequenze automatiche e altre funzionalità del sistema di analisi (*Concordance*, *Frequency* ecc.), ma i dati vengono poi analizzati da un punto di vista qualitativo.

<sup>42</sup> Cfr. [https://www.moreincommon.de/media/leapg0va/more\\_in\\_common\\_studie\\_klima\\_zusammenhalt.pdf](https://www.moreincommon.de/media/leapg0va/more_in_common_studie_klima_zusammenhalt.pdf) [15.06.2023].

Per rispondere alla due domande di ricerca sono state dapprima estratte tutte le parole composte che presentano *Klima* come modificatore. La ricerca<sup>43</sup> ha prodotto 3.497 occorrenze. Queste sono state poi analizzate automaticamente tramite la funzionalità *frequency*, estrapolando così 324 forme lemmatizzate. Grazie a una fine analisi qualitativa le occorrenze sono state ridotte a 2583 e le forme a 278.

Nella Tabella 1.9 sono elencate tutte le parole composte con una frequenza assoluta  $\geq$  a 15:

Lemma	Frequenza assoluta
Klimawandel	924
Klimaänderung	622
Klimaschutz	128
Klimasystem	123
Klimamodell	103
Klimaerwärmung	49
Klimaveränderung	45
Klimaforscher	35
Klimakrise	34
Klimasensitivität	33
Klimazone	32
Klimaschutzmaßnahme	31
Klimaforschung	30
Klimagas	29
Klimapolitik	28
Klimaabkommen	24
Klimaprojektion	23
Klimageschichte	22
Klimaschwankung	21
Klimafaktor	21
Klimawirkung	17
Klimafolgenforschung	16
Klimaneutralität	15
Klimaziel	15

**Tabella 1.9:** Composti con *Klima* come modificatore (frequenza  $\geq$  15)

È possibile notare che quasi tutti i composti sono formati da due unità (tutti sostantivi), eccetto *Klimaschutzmaßnahme* e *Klimafolgenforschung*, formati da tre unità. Le parole composte non hanno alcun elemento di raccordo.

<sup>43</sup> La stringa di ricerca utilizzata è *Klima\**.

*Klimaschutzmaßnahme* e *Klimafolgenforschung* sono casi interessanti anche dal punto di vista semantico, dal momento che il primo non è presente né nel dizionario DWDS<sup>44</sup> né nel dizionario *Duden* online<sup>45</sup>, mentre il secondo è elencato solo nel *Duden* online; questo suggerisce che si tratta di due lessemi piuttosto recenti. Secondo la DWDS-*Wortverlaufskurve* sono, infatti, termini utilizzati dal 1946 ca. in poi. Inoltre, la maggior parte delle parole composte più frequenti è neutra dal punto di vista del significato, ad eccezione di *Klimakrise*, *Klimasensitivität*, *Klimaneutralität*, dalle quali emerge un punto di vista o una valutazione del fenomeno.

Analizzando gli *hapax legomena* (150) è possibile osservare una maggiore creatività: un terzo è formato da tre costituenti (49 su 150) e ben 17<sup>46</sup> presentano un trattino come elemento di raccordo, indice del processo di lessicalizzazione (cfr. Duden 2022: 612). Tra i costituenti vi sono sia elementi stranieri, prevalentemente inglesi, come in *Klimamonitoring*, *Klimakiller*, *Klimaproxies/Klimaproxyis*, *Klima-Thinktank*, *Klimabooklet*, sia nomi propri, come in *Klima-Armageddon* (1), dove si fa riferimento all'Apocalisse, e nell'ironico *Klimagretel*, dove il termine è un appellativo attribuito a Greta Thunberg (2).

(1) Die Menschheit wird aussterben, sollten wir im Juni noch fossile Brennstoffe verwenden. Wenn wir uns bis Juni 2023 nicht von fossilen Brennstoffen trennen werden, bricht die Endzeit aus. Das hatte Thunberg vor bald fünf Jahren, am 21. Juni 2018, auf Twitter vorhergesagt. Bis zum **Klima-Armageddon**, der die gesamte Menschheit auslöschen wird, bleiben uns damit aktuell nur noch fünf Wochen.

(2) Und Ursula van der Leyen hat sie auch noch empfangen! Alles Kasperln im Welttheater. Die **Klimagretel**, Ikone für alle, die ihren Namen klatschen können.

Tra i costituenti vi sono anche acronimi, come in *Klima-FAQ*.

Molti degli *hapax legomena* esprimono una chiara valutazione e presa di posizione, anche con metafore appartenenti a vari ambiti, come per esempio quello medico in *Klimavirus* (3).

(3) Habe gerade ein aktuelles Bild aus der Atacama Wüste bekommen- was steht da inmitten der Einöde? Windräder. Auch die Chilenen sind mit dem **Klimavirus** infiziert.

Il post nell'esempio (3) commenta sarcasticamente un'immagine appena ricevuta del deserto di Atacama, nella quale si intravedono i mulini a vento. Segue

44 Cfr. <https://www.dwds.de> [30.08.23].

45 Cfr. <https://www.duden.de> [30.08.23].

46 Cfr. *Klima-Anpassung*, *Klima-Prognose*, *Klima-Armageddon*, *Klima-Szenarium*, *Klima-FAQ*, *Klima-Variabilität*, *Klima-Forschungszentrum*, *Klima-Thinktank*, *Klima-Tipp*, *Klima-Protest*, *Klima-Idee*, *Klima-Untergang*, *Klima-Aktivistin*, *Klima-Schwurbeln*, *Klima-Daten*, *Klima-Neutralität*, *Klima-Bilanz*.

la valutazione della constatazione che anche i Cileni sono stati infettati dal virus del clima. Sia l'uso del sostantivo composto *Klimavirus* sia del verbo *infizieren* attivano questo campo metaforico.

Nel corpus *Klima\_de* emerge anche la posizione di coloro che sono contrari o critici verso i movimenti in difesa del clima, come in *Klimadiktatur* (4):

(4) Bis auf ganz wenige Ausnahmen geht es den Klimaaktivisten um Macht und Geld und um ein Ziel: Die Transformation zu einer **Klimadiktatur** (Kommunismus unter einer grünen Fahne).

Secondo l'esempio (4) gli obiettivi degli attivisti in difesa del clima sono, infatti, il potere, i soldi e l'instaurazione di una dittatura climatica. Inoltre, emerge chiaramente la posizione di coloro che criticano chi rinnega i cambiamenti climatici ed è ostile agli attivisti, come in *Klima-Schwurblern* (5), *Klimaleugnerbewegung* (6), *Klimawandel-Skeptiker* (7), *Klimawandelleugnung* (8), *Klimakleaktionen*, *Klimakleber*<sup>47</sup> (9).

(5) Wie viele Tonnen CO2 wohl die Greta mit ihren ganzen Reisen schon produziert hat? Naja, zumindest hat sich die Thunfisch-Gretl an den **Klima-Schwurblern** dumm und dämlich verdient!

(6) Zu den wichtigsten Kräften der organisierten **Klimaleugnerbewegung**, die die Existenz der menschengemachten globalen Erwärmung durch gezielte Attacken auf die Klimaforschung abstreiten, zählen das Cato Institute, das Competitive Enterprise Institute, das George C. Marshall Institute sowie das Heartland Institute, allesamt konservativ ausgerichtete Think Tanks.

(7) Auf der Seite **Klimawandel-Skeptiker** setzen wir uns zunächst grundsätzlich mit Klimawandel-Skepsis auseinander und nehmen auf der Unterseite Antworten des UBA auf populäre skeptische Argumente skeptische Thesen genauer unter die Lupe.

(8) Neben individuellen Faktoren wurde – im Kontext der Kontroverse um die globale Erwärmung – u. a. von Naomi Oreskes näher untersucht, wie vor allem durch wirtschaftlich motivierte Interessengruppen gezielt Zweifel an den wissenschaftlichen Erkenntnissen verbreitet wird (vgl. **Klimawandelleugnung**) und wie sich dies wiederum auf politische Entscheidungen auswirkt.

(9) Der Verzicht auf verschlüsselungsintensive Kommunikation, die allein der Unterhaltung dient, wäre ein ausgezeichnetes Opfer für das Klima. Aber das täte ja weh. **Klimakleaktionen** hingegen machen so viel Spaß wie Indianerspiele. Bis auf den polemischen Seitenhieb am Ende könnten das sicher viele unterschreiben. Womöglich sogar die **Klimakleber**. So erfreulich es ist, dass diese Menschen nicht mehr um ihr Leben fürchten müssen, so bitter ist die Lage in dem nordostafrikanischen Staat: In der Hauptstadt Khartum toben weiter Kämp-

47 La *Letzte Generation* (LG) è un gruppo di attivisti che, con azioni provocatorie, ambisce a creare cambiamenti a favore del clima nelle misure governative. Questi attivisti vengono chiamati anche in maniera dispregiativa *Klimakleber*. Cfr. [https://de.wikipedia.org/wiki/Letzte\\_Generation#:~:text=In%20der%20Öffentlichkeit%20stoßen%20diese,straftrechtlichen%20Verurteilungen%2C%20insbesondere%20wegen%20Nötigung](https://de.wikipedia.org/wiki/Letzte_Generation#:~:text=In%20der%20Öffentlichkeit%20stoßen%20diese,straftrechtlichen%20Verurteilungen%2C%20insbesondere%20wegen%20Nötigung) [15.06.2023].

fe, nach Angaben der Vereinten Nationen sind schon mindestens 427 Menschen getötet und 3.700 verletzt worden.

Molti di questi nuovi composti presentano una certa somiglianza con quelli creati durante la pandemia covid (cfr. Flinz 2022, Flinz/Mell 2023).

Dai risultati discussi si può dunque concludere che il discorso del cambiamento climatico analizzato nel corpus CMC costruito *ad hoc* è ricco di composti, usati per esprimere una valutazione o per prendere posizione. Molti di queste parole composte non sono solo semplici occasionalismi, ma possono essere considerati neologismi, dal momento che sono entrati nel lessico della lingua tedesca.

### 1.2.2 Caso studio II: il confisso *-gate* nel corpus *Wikipedia-de*

La *Dudengrammatik* (cfr. Duden 2022: 621) classifica *-gate*<sup>48</sup> come un confisso, dove per confisso si intende «eine weitere Klasse gebundener Morpheme. Konfixe [...], die aus anderen Sprachen – überwiegend Latein und Griechisch, seltener Englisch entlehnt sind».

Di norma i confissi hanno un significato lessicale e si uniscono a una base (*Dieseltgate*) o a un altro confisso (*Pseudo-Gate*) per formare una parola composta. Possono essere anche derivati, ma solo per mezzo di suffissi (cfr. Duden 2022: 621). Alcuni confissi possono essere usati solo come modificatori (in questo caso si tratta di preconfissi), alcuni solo come testa (postconfissi) e altri in entrambi i ruoli: *-gate* può usato solo come postconfisso.

Il confisso *-gate* proviene dall'inglese, ed è considerato un confisso onimico molto produttivo per il tedesco (cfr. Flach/Kopf/Stefanowitsch 2018): si combina preferibilmente con nomi propri e appellativi per riferirsi a “scandali” (cfr. Flach/Kopf/Stefanowitsch 2018: 2). Il confisso è produttivo anche in inglese e presenta attestazioni anche in molte altre lingue (serbo croato, greco, danese, polacco), ma la sua produttività è stata analizzata solo per il cinese (cfr. Cao 2011), l'olandese (cfr. Hüning 2000), l'inglese (cfr. Flach/Kopf/Stefanowitsch 2018) e il tedesco (cfr. Flach/Kopf/Stefanowitsch 2018). Gli studi sull'inglese e sul tedesco lo mettono a confronto con *-skandal* e *-affäre*, sottolineando il suo cambiamento di significato: mentre in passato si usava solo per indicare seri scandali

---

48 In letteratura vi sono varie classificazioni di *-gate*: come suffisso (cfr. Joseph 1998), confisso (cfr. Warren 1990) creato per analogia, contaminazione (cfr. Booij 2010, Kemmer 2003). In questo caso studio *-gate* è considerato un confisso, dal momento che si combina con altri elementi per formare una parola composta, non può stare da solo e ha un significato ben definito (cfr. Donalies 2002, 2009; Fleischer/Barz 2012: 64). Il fatto che si trovino attestazioni di *Gate* anche come morfema libero non contraddice il suo status di confisso (cfr. Flach/Kopf/Stefanowitsch 2018: 9).

politici, oggi *-gate* viene usato anche per eventi scandalosi di basso livello<sup>49</sup>. Il confisso *-gate* si attesta in lingua tedesca negli anni '70 per indicare lo scandalo Barmat, avvenuto durante la Repubblica di Weimar. Negli anni '70-'80-'90 vi sono altre sporadiche attestazioni come in *Watergate*, *Koreagate*, *Dokumenten-Gate*, *Pseudo-Gate*, *Irangate*, *Contragate*, *Waterkantgate*, *Sesselgate*, *Urangate*, *Genscher-Gate*, *Koppigate* e *Nersinggate* (cfr. Flach/Kopf/Stefanowitsch 2018: 7). Attestato è anche l'uso di *Gate* come morfema libero, usato come base per altre modificazioni (cfr. Flach/Kopf/Stefanowitsch 2018: 7).

Nello studio di Flach/Kopf/Stefanowitsch 2018, svolto su corpora di tedesco scritto (DeReKo e il corpus *Zeit* del DWDS), *-gate* si combina principalmente con lessemi collegati allo scandalo in questione: nomi propri (di persona, cognomi, toponimi ecc.), appellativi (parti del corpo, generi alimentari, oggetti, nomi astratti) o altre classi di parola (basi verbali). Nel corso degli anni ci sono state varie fasi, con la creazione di molti nuovi composti che, nella maggior parte dei casi, hanno avuto vita breve (ca. 1 anno). Solo un 6% è rimasto stabile nel corso degli ultimi dieci anni (*Camillagate*). Nel nuovo millennio si è assistito, a differenza dell'inglese, a un nuovo incremento dell'uso del confisso in tedesco (cfr. Flach/Kopf/Stefanowitsch 2018: 16), ma con un cambiamento nella semantica: vengono nominati scandali lievi, di minore entità in vari ambiti (politica, società ecc.) (cfr. Flach/Kopf/Stefanowitsch 2018: 19). Lo shift è da scandali politici a scandali di minore entità, che riguardano la vita quotidiana: mentre la combinazione con nomi propri è indice di un maggior scandalo (*Pauli-Gate*), l'uso di appellativi è indice di minor scandalo (*Döner-Gate*) (cfr. Flach/Kopf/Stefanowitsch 2018: 22).

Obiettivo del presente caso studio è quello di analizzare il confisso *-gate*<sup>50</sup> in un corpus dell'archivio *WP – Wikipedia Artikel u. Artikel- & Benutzerdiskussionen 2013/15/17/19* di DeReKo: il corpus *ndd19 – alle Wikipedia-Diskussionen zu Artikeln bis 2019*<sup>51</sup> (Tabella 1.10).

Corpus	Testi	Parole	Data
<i>ndd19</i>	711.935	415.929.118	2002-2019

**Tabella 1.10:** Caratteristiche del corpus *ndd19*

Dal corpus sono state estratte 713 forme, ossia 4770 risultati. Grazie a una fine analisi qualitativa le forme sono state poi ridotte a 134: sono state eliminate tutte quelle non pertinenti o errate, come per esempio *congregate* (verbo), *Barthgate*

49 Anche in italiano il confisso è produttivo. Attualmente la discussione nei media ruota intorno al cosiddetto “pandoro-gate”, in riferimento allo scandalo che ha coinvolto Chiara Ferragni e l'azienda Balocco.

50 La ricerca è stata effettuata con la stringa: \*gate.

51 Il corpus *ndd19* raccoglie le pagine di discussione sui contenuti degli articoli di *Wikipedia*.

(nome proprio di luogo), *Ashgate* (nome proprio di persona), *Researchgate* (nome proprio di una piattaforma).

Anche nel corpus *mdd19* emergono composti con nomi propri di persona, di luogo, di prodotti ecc.

Beetlegate, Bengasi-Gate/Bengasigate, Chinagate, Ibizagate, Irangate, Katrina-gate, Köln-Gate, Koreagate, L'Angolagate, Maddiegate, Maggiegate, Merkelgate, Monicagate, Moskau-Gate, Qatargate, Rolexgate, Rubygate, Russia-Gate/Russia-gate/Russigate, Schulz-Gate, Selmayrgate, Spiegel-Gate/Spiegelgate, Teletubbiegate, Trump-Gate/Trumpgate, Washingtonegate, WATERGATE/Watergate/watergate, Waterkantgate ecc.

Dalle parole composte si evincono chiaramente le persone, i luoghi, i prodotti coinvolti nello scandalo politico o nell'evento in questione. Vi sono scandali di varia entità, legati alla politica internazionale, come il *Bengasi-Gate*, lo scandalo che ha coinvolto Hillary Clinton, accusata di aver continuato a usare, anche per comunicazioni ufficiali, la posta elettronica personale; scandali nella politica nazionale tedesca, come il *Waterkantgate*, noto anche come la *Kieler Affäre* in cui era coinvolto il politico Uwe Barschel, o l'*Ibizagate*, conosciuto anche come *Strache-Affäre*, in cui protagonista dello scandalo era il politico austriaco Heinz-Christian Strache. Ma vi sono anche scandali legati a prodotti, come il *Beetlegate*, dove *Beetle* è il nome di un'auto della Volkswagen (lo scandalo è più comunemente conosciuto come *Diesel-Gate*), a persone di cui si è parlato nella cronaca per la loro scomparsa (*Maddiegate*) ecc. O semplicemente a personaggi o eventi per il quale deve essere redatto un articolo in *Wikipedia*: *Teletubbiegate*. Con *Teletubbiegate* si intende, infatti, la discussione sul lemma *Teletubbie* in *Wikipedia* (per la funzione specifica di altri confissi in *Wikipedia* cfr. Gredel/Flinz 2020; Flinz/Gredel 2022).

Nel corpus sono presenti anche composti con appellativi (parti del corpo, generi alimentari, oggetti, nomi astratti), che indicano l'oggetto della discussione/scandalo:

Clima-Gate/Climagate/climagate, Climate-Gate/Climate-gate/climate-gate/CLIMATEGATE/ClimateGate/Climategate/climategate, ClimatGate/Climatgate, Klima-Gate/Klimagate, Diesel-Gate/Dieslegate, Dirndlgate, E-Mail-Gate, Fischfuttergate/fischfuttergate, Gaucho-Gate/Gauchogate, Handtaschen-Gate, Handygate, Nippelgate, Nippelgate, PizzaGate/Pizzagate/pizzagate, Queensgate, Steuer-Gate/Steuergate, Toilettengate ecc.

Le parole composte denotano sia eventi di portata internazionale, come le discussioni sul clima, sul diesel, sulle tasse, ma anche scandali politici di varia entità, come lo scandalo sul complotto *Pizzagate* nato durante le elezioni presidenziali statunitensi del 2016, così come scandali più lievi, come il *Toilettengate*

intorno al politico Gregor Gysi del partito *Die Linke*, che fu seguito e filmato fino al bagno. Oppure eventi mediatici come il *Gaucht-Gate*/*Gauchogate* della nazionale di calcio tedesca o il *Nippelgate* intorno all'esibizione di Janet Jackson e Justin Timberlake al 38 Super Bowl del 1. Febbraio 2004.

Il cambiamento semantico è visibile anche in questo corpus: vengono nominati scandali lievi, di minore entità in vari ambiti (politica, società ecc.). Non tutti i nomi propri sono collegati all'ambito politico (*Maddiegate*) e anche negli appellativi vi è un riferimento a scandali politici di minore entità (*Toilettengate*). Da un punto di vista testuale-pragmatico vi sono esempi in cui è evidente la funzione della tematizzazione metacomunicativa (ossia metalinguistica) di aspetti relativi alla produzione testuale collaborativa, come in *Citationgate* (1) o in *Zitate-Gate* (2):

- (1) Mein einsetzbarer Zeitaufwand z.Zt. hält sich in engen Grenzen und zudem bin ich mit den Nachbeben von  
<http://mail.wikimedia.org/pipermail/wikien-l/2006-September/054633.html>  
**Citationgate** beschäftigt. Aber in Grenzen mache ich gerne mit. (wdd19)
- (2) P. S.: Und noch so ein merkwürdiger, intransparenter Terminus: „Zusammengooglen“. Klingt mir nach dem internetfeindlichen Populismus eines Sascha Lobo, der mit dem Medium Internet grundsätzlich überfordert ist, wie man auch an seiner SPIEGEL-Wortmeldung zum gegenwärtigen: „**Zitate-Gate**“ sehen kann. Was aus diesem ominösen Internet oder von Google kommt, muß ja von vornherein schon falsch sein, auf einer solchen diffusen Emotion reitet das Wort vom: „Zusammengooglen“ und seine etymologischen Verwandten. Was sich nicht nur auf die, wie von Lobo attestiert, angeblich: „zusammengetwitterten“ und: „verkürzten“ Zitate bezieht. (wdd19)

Dal punto di vista morfologico è osservabile la presenza di un alto numero di possibilità di combinazione, con varianti di scrittura, costituenti stranieri (*Climagate*, *Climategate*, *Pizzagate*, *Gauchogate*, *Queensgate*) o autoctoni (*Klimagate*, *Dirndlgate*, *Fischfuttergate*, *Handtaschengate*), con abbreviazioni (*BfR-Gate*, *FAZ-Gate*), con o senza trattino di raccordo (*Gaucht-Gate*, *Gauchogate*), con l'uso di sole maiuscole (*CLIMATEGATE*), con l'uso di sole minuscole (*climagate*), con l'uso di lettere maiuscole e minuscole (*ClimaGate*). L'elemento di raccordo è presente raramente (*Queensgate*). Sono state individuate anche costruzioni creative come *ErwärmungKlimagate*, *Faz.net-Gate*, *Iran-Contra-Gate*, *Nixon/Watergate*, *StopGamerGate* ecc. Si tratta di costruzioni create *ad hoc*, tipiche della scrittura interattiva (cfr. Storrer 2013: 337), legate a situazioni specifiche e necessarie di interpretazione.

Come è stato mostrato da questo caso studio, l'uso del confisso *-gate* non è identico in tutti i corpora e le caratteristiche del corpus devono necessariamente confluire nella discussione dei risultati.

## 1.3 Attività ed esercizi<sup>52</sup>

### 1.3.1 La parola e i suoi elementi costitutivi

Teoria → Capitolo 1.1.1

1. Estrarre dal corpus *German Web 2020* per mezzo della funzionalità *wordlist* di *Sketch Engine* le seguenti liste di frequenza: 1. *wordlist* (opzione *word*) e 2. *wordlist* (opzione *lemma*). Confrontare le liste e discutere i risultati.
2. Estrarre le occorrenze di un verbo a piacere in un corpus a scelta. Analizzare 10 occorrenze e classificare il verbo scelto in base alle sue caratteristiche principali (persona, numero, tempo e modo). Fare la stessa cosa con un sostantivo, con un aggettivo, con un articolo e con un pronome: in questo caso analizzare genere, numero e caso.
3. Cercare un avverbio, una preposizione, una congiunzione e una particella in un corpus a scelta e verificarne l'invariabilità.
4. Estrarre la lista di frequenza lemmatizzata (*Wordlist* → *lemmas*) di un corpus in *Sketch Engine*. Determinare se si tratta di parole con significato pieno e parole con significato grammaticale-relazionale. Eseguire lo stesso procedimento con la lista di frequenza di parole (*Wordlist* → *words*).
5. Consultare i significati della parola *Brücke* in un dizionario monolingue. Estrarre 50 KWIC di *Brücke* in un corpus e riflettere sui significati.
6. Estrarre tramite la ricerca *simple* la sequenza \*st. Per prima cosa indicare quali delle parole estratte sono verbi. Poi individuare per ciascun verbo la radice e il morfema per la 2. persona singolare.
7. Cercare in un corpus tutte le parole composte di *Sommer* (Sommer\*). Identificare i sostantivi e analizzarli distinguendo tra morfemi lessicali e grammaticali: *Sommerausrüstung*, *Sommerabenteuer*, *Sommerferien*, *Sommerfest*, *Sommerpause*, *Sommersaison*, *Sommersemester*.
8. Verificare che *Schwieger* non è un morfema libero. Sicuramente ci saranno casi bizzarri, come *Peter Schwieger*: di cosa si tratta? Discuterne. Successivamente cercare *Schwieger\**. Riflettere sui risultati.
9. Cercare \*fremd\*. Discutere i risultati: es. *entfremden*, *Fremdenverkehr*, *fremden*, *fremder*, *Fremdwerbung*, *fremde*, *Entfremdung*, *entfremden*, *Fremdentwicklungen*, *Fremdanbietern*, *Fremde*, *Fremdsein*, *Fremdkörper*.
10. Cercare i lessemi che iniziano con il prefisso {un} (ricerca con un\*) o con il prefisso {ver} (ricerca con ver\*). Discutere i risultati (5 per tipo).
11. Cercare quali parole si formano con i suffissi {heit} (ricerca con \*heit), {lich} (ricerca con \*lich), {thek} (ricerca con \*thek). Discutere i risultati (5 per tipo).
12. Cercare ge\*t (di cosa si tratta?). Identificare tra le forme estratte 5 participi passati e dire da quale forma base provengono.

<sup>52</sup> Per la scelta della forma linguistica “a piacere” usare come supporto la parte teorica.

13. Cercare le occorrenze di *untüchtig*. Analizzare 5 KWIC. Estrarre ora \**untüchtig*. Discutere i risultati.
14. Cercare \**un* e \**in* e riflettere sui risultati. Sono la negazione di quale parola? Fare poi la ricerca con \**anti*. Le occorrenze estratte indicano un'opposizione a cosa? Cercare \**vor*: le occorrenze indicano anteriorità a cosa? Cercare \**er*: le parole estratte indicano una persona (colui che compie un'azione) o uno strumento?

### 1.3.2 Costruzione delle parole: la flessione

Teoria → Capitolo 1.1.2.1

1. Cercare 5 neologismi<sup>53</sup> in un corpus e discutere i risultati.
2. Cercare i seguenti sostantivi in un corpus e vedere se ci sono delle varianti: *Pizza*, *Spaghetti*, *Lockdown*, *Shutdown*. Riflettere anche sul plurale e sul genitivo, dopo aver fatto le apposite ricerche.
3. Cercare i seguenti verbi in un corpus e vedere se ci sono delle varianti: *googeln*, *downloaden*. Riflettere anche sul participio passato.
4. Cercare il suffisso {*ling*} e verificare che i sostantivi estratti siano di genere maschile (stringa di ricerca: \**ling*, estrarre un *sample* di 200, orientare le KWIC sul contesto di sinistra e focalizzare solo le parole che finiscono per *-ling* e non per esempio *-linge* ecc.). Fare la medesima ricerca con i suffissi {*keit*} e {*lein*}. Che cosa si può dedurre?
5. Identificare i sostantivi che finiscono in *-e* (cfr. *schwa* come in *Tante*). Di che genere sono? (cercare con la stringa \**e*)
6. Identificare i sostantivi che finiscono in *-en*. Di che genere sono? (= cercare con la stringa \**en*)
7. Cercare il genere delle seguenti parole *Korpus* (stringhe di ricerca: *der Korpus* e *das Korpus*) e *Pub* (stringhe di ricerca: *der Pub* e *das Pub*). Quale genere è più frequente? Se la ricerca viene fatta in corpora di diverso tipo (*German Web*, *DeReKo*, *FOLK*) emergono differenze?
8. Cercare le occorrenze con *-keit*. Identificare le forme al plurale. Quale morfema è stato utilizzato per fare il plurale? Cercare le occorrenze con *-ling*. Identificare le forme al plurale. Quale morfema è stato utilizzato per fare il plurale? Cercare le occorrenze con *-tum*. Identificare le forme al plurale. Quale morfema è stato utilizzato per fare il plurale? Cercare le occorrenze con *-chen*. Quale morfema è stato utilizzato?
9. Riflettere sulla declinazione dell'aggettivo. Usando le seguenti stringhe di ricerca (*der schön\* Anzug*; *der blau\* Anzug*; *der neu\* Anzug*) riflettere sui risultati. Fare poi la ricerca con *schön\*Anzug*; *blau\* Anzug*; *neu\* Anzug*. E

53 Per la scelta di neologismi cfr. il *Neologismenwörterbuch* del *Leibniz-Institut für Deutsche Sprache* (<https://www.owid.de/docs/neo/start.jsp> [25.07.23]) o la banca dati *Wortwarte* <https://wortwarte.de> [25.07.23].

successivamente con ein schön\* Anzug; ein blau\* Anzug; ein neu\* Anzug. Discutere i risultati.

10. Cercare le forme dei verbi al passato (stringa di ricerca: [tag="VFIN.\*Past.\*"] in CQL) e riflettere se si tratta di un coniugazione debole o forte.

### 1.3.3 Costruzione delle parole: la formazione delle parole

Teoria → Capitolo 1.1.2.2

1. Verificare le occorrenze di *Him*. La parola è presente nel corpus? Che cosa si può affermare dall'analisi delle KWIC? È *Him* una parola tedesca a sé stante?
2. Cercare nel corpus *Italian Web 2020* le parole composte con *capo* (stringa di ricerca: capo\*; *get a random sample* di 200 KWIC; orientare le occorrenze in base alla KWIC). Che cosa può essere dedotto dai risultati della ricerca? (*capobanda* è il capo della banda; *capocomico* è il capo di chi/di che cosa?). Quale è la testa della parola composta? Quale è il modificatore della parola composta? Quale è quindi l'ordine della parola composta?
3. Cercare nel corpus *German Web 2020* le parole composte con *Schlüssel* (stringa di ricerca: \*schlüssel; *get a random sample* di 200 KWIC; orientare le occorrenze in base alla KWIC). Che cosa può essere dedotto dai risultati della ricerca? (*Autoschlüssel* è la chiave di chi/di che cosa? *Briefkastenschlüssel* è la chiave di chi/di che cosa?). Quale è la testa? Quale è il modificatore della parola composta? Quale è quindi l'ordine dei costituenti nella parola composta?
4. Cercare le parole composte di *Flasche* (stringa di ricerca: \*flasche; *get a random sample* di 200 KWIC; orientare le occorrenze in base alla KWIC). Identificare il genere della parola composta da N+N. Scomporre la parola composta nei due sostantivi (N) e identificare i rispettivi generi. Che cosa può essere concluso? Ora indicare il plurale della parola composta e dei singoli sostantivi. Che cosa può essere dedotto?
5. Cercare le occorrenze di *süßsauer* e di *sauersüß*. Quale delle due parole è più frequente e quindi la più usata? Cercare le occorrenze di *rotgrün* e di *grünrot*. Quale delle due parole è più frequente e quindi la più usata?
6. Identificare la testa delle seguenti parole composte: *Autoschlüssel*, *süßsauer*, *freibalten*, *danksagen*, *überallhin*, *übermorgen*, *Bierflasche*, *hilfsbereit*. Classificarle in composto nominale, composto verbale, composto aggettivale, composto avverbale.
7. Cercare le parole composte di *Schuh* (stringa di ricerca: \*schuh; *get a random sample* di 200 KWIC; orientare le occorrenze in base alla KWIC). Dire che cosa indica il modificatore (materiale? luogo? provenienza? causa? strumento? ...?)

8. Cercare le parole composte di *Lehrer* (stringa di ricerca: \*lehrer; *get a random sample* di 200 KWIC; orientare le occorrenze in base alla KWIC). Che cosa indicano questi composti di reggenza? *Mathelehrer* = *jemand, der Mathe lehrt*; *Erdkundelehrer* = *jemand, der Erdkunde lehrt*; ecc.)
9. Cercare le parole composte di *Kiste* (stringa di ricerca: \*kiste; *get a random sample* di 200 KWIC; orientare le occorrenze in base alla KWIC). Analizzare le parole composte dal punto di vista del significato: *Bananenkiste*, *Bastelkiste*, *Blechkiste*, *Fischkiste*, *Holz-kiste*? Dire che cosa indica il modificatore (materiale? luogo? provenienza? causa? strumento? ...?) Ci sono parole composte che possono avere più interpretazioni?
10. Cercare in un corpus a piacere le possibilità di raccordo nel composto formato da *Asche* (modificatore) e *Wolke* (testa) (usa la stringa di ricerca: asche\*wolke; *get a random sample* di 200 KWIC; orientare le occorrenze in base alla KWIC). Quali varianti sono state riscontrate? Quale è la più frequente? Fare ora la ricerca con *Schwein* e *Braten*. Quali varianti sono state riscontrate? Quale è la più frequente?
11. Cercare in un corpus a piacere le possibilità di raccordo nel composto creativo formato da *Bahn* e *Card* (stringa di ricerca: Bahn\*Card; *get a random sample* di 200 KWIC; orientare le occorrenze in base alla KWIC). Eseguire anche la ricerca con i due elementi separati da uno spazio. Vi sono occorrenze? Quale è la variante più frequente. Eseguire la medesima procedura con *Kunden Service*.
12. Cercare in vari corpora monolingui del tedesco i composti delle seguenti parole usate come modificatore: *Kinder*, *Eltern*, *Krieg*, *Auto*, *Reise*. Che cosa si può osservare?
13. Riflettere sulla coppia minima *kindisch* – *kindlich*. Per prima cosa consultare il significato in un dizionario monolingue (Duden, DWDS ecc.). Poi cercare in un corpus attestazioni che mettano in evidenza la differenza.
14. Suddividere in base e suffisso le seguenti parole: *Lehrer*, *amtieren*, *höflich*, *kreuzweise*. Classificare poi il tipo di suffisso in nominale, verbale, aggettivale e verbale. Cercare poi le parole in un corpus e verificare la classe di parola della parola derivata. Che cosa può essere dedotto?
15. Partendo dal presupposto che anche nella derivazione domina il principio della testa a destra, verificare in un corpus il genere dei seguenti sostantivi: *Lehrer* (*Lehr-er*), *Bohrer* (*Bohr-er*), *Seufzer* (*Seufz-er*). Indicare da quale verbo provengono. Riflettere su cosa potrebbe indicare il sostantivo (la persona che svolge l'azione? lo strumento? l'azione stessa?). Cercare altri sostantivi che finiscono in *-er* ed eseguire le medesime riflessioni.
16. Cercare in un corpus cinque sostantivi che finiscono in *-ung*. Verificare che siano femminili e identificare se si tratta del processo o del risultato di un'azione.

17. I suffissi {chen} e {lein} esprimono una riduzione della dimensione, anche in senso affettivo. La riduzione può essere positiva, come in *Kätzchen*, o negativa con una punta di disprezzo come in *Büirschchen*. Trovare in un corpus 10 parole che finiscono in *-chen* e in *-lein*, classificarle, usando anche il supporto di un dizionario (Duden, DWDS) in positive e negative.
18. Cercare in un corpus aggettivi che finiscono in *-bar* (stringa di ricerca: \*bar). Determinare da quale verbo sono derivati.
19. Cercare in un corpus le parole che contengono *-ruhe* (stringa di ricerca: \*ruhe; *get a random sample* di 200 KWIC; orientare le occorrenze in base alla KWIC). Identificare almeno 2 parole derivate dal sostantivo *Ruhe*.
20. Cercare in un corpus sostantivi con il suffissoide *-bombe*. Sceglierne 5 e analizzare il loro significato, aiutandosi con le KWIC e con dizionari monolingue (Duden, DWDS).
21. *Himi* è l'acronimo di *Hilfswissenschaftler* e EDV di *elektronische Datenverarbeitung*. Cercare in un corpus le occorrenze degli acronimi e dopo aver orientato il contesto a sinistra, identificare il genere. Qual è il plurale di *Himi*?



# Capitolo 2

## Sintassi

Oggetto di questo capitolo è la sintassi, l'ambito della linguistica che si occupa della combinazione delle parole per la costruzione della frase (cfr. Pittner 2016: 86). La sintassi, infatti, descrive «die Regularitäten, nach denen Wörter zu Sätzen verknüpft werden» (Duden 2022: 25), ossia le regolarità per le quali<sup>1</sup> le parole si uniscono in unità più grandi come il costituente o sintagma (*Phrase*), la frase (*Satz*) e il periodo (*Satzgefüge*).

La parola è l'unità più piccola della sintassi, mentre la frase è quella più grande. Ma che cosa è una frase? Dare una definizione univoca non è facile, dal momento che in base al punto di vista teorico adottato esistono varie possibilità. Di Meola (2014: 91) intende per frase una costruzione autonoma che non fa parte di costruzioni più grandi<sup>2</sup>. La *Dudengrammatik* (cfr. Duden 2022: 27-32) specifica che una frase deve contenere un verbo finito. Bianco/Mollica (2024) aggiungono l'informazione che oltre al predicato la frase contiene gli elementi selezionati direttamente dal verbo, i complementi di verbo (detti anche attanti, in tedesco *Ergänzungen*), ed eventualmente anche gli elementi non dipendenti dal verbo, i circostanziali (detti anche complementi di frase, in tedesco *Angaben*). Si osservi la seguente frase (1):

(1) Der Gewinner des Turniers fährt nach Japan. (deTenTen20)

La frase in (1) è formata da un predicato (*fährt*) che seleziona un soggetto (*der Gewinner des Turniers*) e un complemento di luogo (*nach Japan*). Sia il soggetto, sia il complemento di luogo sono richiesti dal verbo e sono quindi i complementi di verbo, le *Ergänzungen*. A questi elementi possono essere aggiunti ulteriori elementi, che non sono richiesti dal verbo, ma che aggiungono ulteriori informazioni, per esempio di tipo temporale, spaziale, modale ecc. come nella frase (2):

- 
- 1 Gaeta (2017: 174) oltre alla funzione della linearizzazione delle parole (*Wortstellung*), aggiunge quella dell'esplicitazione delle relazioni tra il piano delle rappresentazioni dei significati delle singole unità che si combinano nel significato complessivo della frase e la loro codificazione lineare (cfr. Gaeta 2017: 174).
  - 2 Gaeta (2017: 171) definisce la frase come «la più piccola unità sintattica autonoma in grado di dare adito a una predicazione», dove per predicazione si intende l'attribuzione per mezzo della funzione  $f(x)$  di una proprietà  $P$  a una variabile  $x$  come nell'esempio da lui discusso (*Niki schläft*), in cui alla variabile *Niki* viene attribuita la proprietà SCHLAF( $x$ ) dal predicato *schlafen*. Oltre al criterio semantico della predicazione la frase deve avere i requisiti di minimalità e indipendenza. Gaeta (2017: 173), postulando un continuum in cui al centro vi sono le frasi prototipiche, confronta i vantaggi di tale definizione con altre utilizzate in letteratura.

(2) Der Gewinner des Turniers fährt 2025 nach Japan.

“2025” è l’elemento che è stato aggiunto e che specifica l’azione da un punto di vista temporale: è il cosiddetto circostanziale (*Angabe*).

La frase, oltre alla forma, deve soddisfare anche i criteri di contenuto, ossia descrivere uno stato di cose (*Sachverhalt* o *Proposition*), e di funzione, ossia avere una funzione comunicativa o pragmatica (con una frase, infatti, viene messa in atto un’azione comunicativa) (cfr. Duden 2022: 27-32). La frase prototipica ha quindi le seguenti caratteristiche: presenta un verbo finito e gli elementi selezionati direttamente dal verbo, soddisfa il criterio di contenuto (descrive uno stato di cose), svolge una funzione comunicativa<sup>3</sup>.

Partendo da una sintetica introduzione teorica, si mostra tramite due casi studio come i corpora possano essere utilizzati per illustrare determinati fenomeni sintattici, per fare ipotesi e per lavorare sia in maniera induttiva che deduttiva. Seguiranno alcune attività che prevedono l’uso di corpora e dei suoi strumenti di analisi.

## 2.1 Introduzione

I parlanti nativi riescono intuitivamente a distinguere se una determinata frase è grammaticale o meno<sup>4</sup>. Sanno, infatti, che per formare frasi grammaticali le parole devono seguire un determinato ordine (*Abfolge*) ed essere flesse correttamente (*morphologische Markierung*). A ciò si aggiunge un terzo aspetto: l’interpunzione nello scritto o l’intonazione nel parlato (cfr. Pittner 2016: 86-87).

Come già accennato, le unità elementari della sintassi sono le parole, che vengono combinate per realizzare costruzioni sintattiche a diverso grado di complessità. Si possono distinguere due tipi di categorie sintattiche: le classi di parola (*Wortarten*, anche *Wortklassen*, *lexikalische Kategorien*)<sup>5</sup> e i sintagmi (*Phrasen*).

### 2.1.1 Classi di parola

Le parole possono essere raggruppate in base criteri morfologici, sintattico-distribuzionali e semantici<sup>6</sup> in classi di parola o parti del discorso (*Wortarten*, anche *Wortklassen*, *lexikalische Kategorien*).

In base al criterio morfologico le parole possono essere categorizzate in parole variabili (*veränderbare* o anche *flektierbare Wörter*) e parole non variabili (*unveränderbare Wörter* o anche *nicht flektierbare Wörter*). Tra le parole variabili si

3 La *Dudengrammatik* (cfr. Duden 2022: 32) aggiunge altri due criteri: il criterio della parentesi frasale (*Satzrahmen*) e quello dell’occupazione del *Nachfeld* (campo posposto).

4 Oltre alla grammaticalità, un parlante nativo può anche capire se la frase è comprensibile e accettabile dal punto di vista stilistico (cfr. Busch/Stenschke 2018: 124-125).

5 Cfr. anche paragrafo 1.1.1.

6 I tre criteri sono in ordine di importanza (cfr. Pittner 2016: 88).

distingue, inoltre, tra parole declinabili (*deklinierbare Wörter*) e parole coniugabili (*konjugierbare Wörter*) (cfr. Tabella 2.1). Le parole declinabili possono essere poi raggruppate in base al criterio sintattico-distribuzionale, dove per distribuzione si intendono le posizioni in cui un elemento può trovarsi all'interno di una frase.

Wort			Wortart
	flektierbar	deklinierbar	Substantiv, Adjektiv, Artikel, Pronomen
		konjugierbar	Verb
unflektierbar		Adverb, Präposition, Konjunktion, Adjunktör, Partikel, Interjektion	

**Tabella 2.1:** Suddivisione delle parole in base al criterio morfologico<sup>7</sup>

Qui di seguito un breve excursus sulle parole variabili (1.) e non variabili (2.):

### 1. Parole variabili:

Le parole variabili comprendono i nomi (*Nomina/Substantive*), gli aggettivi (*Adjektive*), gli articoli (*Artikel o Artikelwörter*) e i pronomi (*Pronomina*).

#### *I nomi o sostantivi*

I nomi (*Nomina*) o sostantivi (*Substantive*) variano in base alla categoria del numero (*Numerus*) – singolare e plurale – e del caso (*Kasus*) – nominativo, genitivo, dativo e accusativo. Sono dotati di un genere fisso (*Genus*), che può essere maschile (*Maskulinum*), femminile (*Femininum*) e neutro (*Neutrum*), e possono essere accompagnati da un articolo<sup>8</sup>.

Da un punto di vista semantico si può distinguere tra (cfr. Duden 2022: 698):

- nomi concreti (*Konkreta*) e nomi astratti (*Abstrakta*), dove i primi si riferiscono a entità concrete (*Mensch, Kind, Fisch, Rose, Wald, Rom, Karl*), mentre i secondi ad azioni, stati, caratteristiche, relazioni, grandezze ecc., come *Schlaf, Reise, Liebe, Meter* ecc. Tra i nomi concreti rientrano anche i nomi collettivi e di massa (*Stoffsubstantive*), come *Mehl, Reis*;
- nomi animati (*belebt*) e nomi inanimati (*unbelebt*). Grazie a questa caratterizzazione possono essere precisati alcuni nomi concreti: *Mensch, Katze, Baum* vs. *Fahrzeug, Weg, Stein*;

<sup>7</sup> Per la tabella in italiano si veda il paragrafo 1.1.1 (PARTE II). Anche la *Dudengrammatik* (cfr. Duden 2022: 594) distingue tra parole variabili (chiamate *flektierbar*) e non variabili (chiamate *nicht flektierbar*).

<sup>8</sup> Alcune precisazioni sono però doverose. Non tutte le categorie di sostantivo ammettono, infatti, un articolo (i nomi collettivi, i nomi astratti e i nomi propri sono in maggioranza senza articolo); alcuni nomi possono avere un genere diverso in base al significato (*der Teil – das Teil*) o alla regione (*der Radio – das Radio*); alcuni sostantivi possono essere solo al singolare (*Sand*), altri solo al plurale (*Geschwister*). Si veda anche il paragrafo 1.1.2.1.

- appellativi (*Appellativa, Gattungsnamen*), come *Hund, Mann, Flasche*, e nomi propri (*Eigennamen*), come *Sarah, Tobias*.

### *Gli aggettivi*

Gli aggettivi hanno come categorie flessive attive il numero, il caso e il genere, ed è la presenza o assenza dell'articolo a determinare la marca flessiva adeguata (cfr. Gaeta 2017: 107). Gli aggettivi presentano forme di intensificazione (*Steigerung*) e possono essere nella maggior parte dei casi usati in forma comparativa (*Komparativ*) e superlativa (*Superlativ*): *schön – schöner – am schönsten*<sup>9</sup>.

Gli aggettivi possono assumere diverse posizioni all'interno della frase<sup>9</sup>. Quando si trovano tra articolo e sostantivo hanno funzione attributiva; quando, insieme all'ausiliare, costituiscono il predicato hanno funzione predicativa. Se usati in funzione attributiva vengono declinati<sup>10</sup> (3), mentre non vengono declinati se usati in funzione predicativa (4):

(3) Sie suchen das **schöne** Leben, die totale Freiheit. (deTenTen20)

(4) Die Bergwelt ist atemberaubend **schön** und faszinierend. (deTenTen20)

Alcuni aggettivi possono essere usati in un'unica funzione: per esempio *gestrig* può essere usato solo in funzione attributiva mentre *quitt* solo in funzione predicativa<sup>11</sup>. Ulteriori funzioni sono quella avverbiale (5) e quella di intensificatore (6). In entrambi i casi gli aggettivi non vengono declinati.

(5) Es klingt alles zu **schön**. (deTenTen20)

(6) Während der Monsunzeit von Juni bis September ist es zwar auch **schön** warm, aber es regnet viel und durch die hohe Luftfeuchtigkeit trocknet es nicht wieder. (deTenTen20)

Dal punto di vista semantico si distingue tra aggettivi qualificativi (*qualifizierende Adjektive*), aggettivi quantificativi (*quantifizierende Adjektive o Zahladjektive*) e aggettivi relazionali (*relationale Adjektive*) (cfr. Duden 2022: 768-769). Gli aggettivi qualificativi identificano una proprietà che può avere una diversa intensità (*lang – kurz*), anche in scala (*kalt – kühl – lau – warm*), anche con carattere valutativo (*gut – böse*). Gli aggettivi quantificativi indicano una quantità o una

9 La definizione presentata nella *Dudengrammatik* (cfr. Duden 2022: 766) «Adjektive sind Lexeme, deren Formen zwischen Artikel und Nomen stehen können» mette in primo piano la posizione.

10 Vi sono eccezioni come, per esempio, alcuni aggettivi usati per esprimere un colore (*Farbadjektive*) che non vengono declinati: *eine prima Idee, das lila Kleid* ecc. (cfr. Duden 2022: 767).

11 In letteratura non vi è accordo nell'inserire questo tipo di aggettivi tra le parole declinabili. Engel (2009) li chiama *Kopulapartikeln* e Zifonun/Hoffmann/Strecker (1997) *Adkopula*. La *Dudengrammatik* (cfr. Duden 2022: 767) constata che in alcune grammatiche sono inseriti anche tra gli avverbi. In questo volume vengono considerati aggettivi.

posizione in una successione (*eins, zwei, drei, erste, zweite, dritte*), mentre quelli relazionali una relazione o un'appartenenza (*afrikanisch, afghanisch; evangelisch, katholisch* ecc.).

### *Gli articoli*

Gli articoli occupano la posizione iniziale dei sintagmi nominali e non sono cumulabili (cfr. Gaeta 2017: 123). Come già accennato, tra un articolo e un sostantivo possono trovarsi anche aggettivi o altro materiale linguistico<sup>12</sup>.

La funzione dell'articolo è quella di determinare la referenza (*Referenz*) del sostantivo:

- se il sostantivo, grazie al contesto o cotesto<sup>13</sup>, è identificato in maniera definita e precisa, l'articolo è determinativo o definito (*bestimmter Artikel*): *der*;
- se invece la grandezza non è definita, l'articolo è indeterminativo o indefinito (*unbestimmter Artikel*): *ein*.

Dall'articolo si può identificare il caso del sostantivo.

Oltre all'articolo determinativo e quello indeterminativo vi sono l'articolo negativo (*kein*), l'articolo possessivo (*mein*), l'articolo dimostrativo (il prossimale *dieser*, il distale *jener*, il topicale *solcher*), l'articolo interrogativo (*welcher*), l'indefinito universale (*jener*) e l'indefinito esistenziale (*mancher, irgendein*)<sup>14</sup> (cfr. Gaeta 2017: 123).

### *I pronomi*

I pronomi<sup>15</sup> sono una classe molto eterogenea sia dal punto di vista formale, sia dal punto di vista funzionale<sup>16</sup>. Si dividono in<sup>17</sup>:

- pronomi personali (*Personalpronomen*): *ich, du, er, sie, es, wir, ihr*<sup>18</sup>;
- pronomi possessivi (*Possessivpronomen*): *mein, dein, sein, ihr, unser, euer*<sup>19</sup>;

12 La struttura a parentesi è attribuita anche ai gruppi nominali (cfr. Weinrich 2005). L'apertura di parentesi è un articolo (determinativo, indeterminativo, dimostrativo, possessivo, un'espressione di quantità), mentre la chiusura è il sostantivo (cfr. Blühdorn/Foschi 2014: 44). Si veda anche la *Dudengrammatik* (cfr. Duden 2022: 402s.; 729).

13 Per "cotesto" si intende il contesto situazionale linguistico interno al testo. Si differenzia dal "contesto", riferito, invece, alle condizioni esterne al testo (cfr. Bußmann 2007: 173).

14 Per una panoramica degli articoli e delle loro funzioni si veda anche la *Dudengrammatik* (cfr. Duden 2022: 730).

15 Alcune grammatiche inseriscono gli articoli e i pronomi in una sola classe: Gaeta (2017: 123) usa il termine "determinante".

16 Cfr. *Grammis* (<https://grammis.ids-mannheim.de/systematische-grammatik/274>). [15.06.2023].

17 Per una panoramica dei pronomi e delle loro funzioni si veda anche la *Dudengrammatik* (cfr. Duden 2022: 730).

18 Le forme dei pronomi personali si distinguono per persona, caso, numero e a volte anche per genere (cfr. Duden 2022: 735).

19 A ogni pronome personale corrisponde un possessivo.

- pronomi dimostrativi (*Demonstrativpronomen*): *der/die/das, dieser, jener, derjenige, derselbe*;
- pronomi negativi (*Negationspronomen*): *kein, nichts*;
- pronomi riflessivi (*Reflexivpronomen*): *sich*<sup>20</sup>;
- pronomi interrogativi (*Fragepronomen*): *wer/was, welcher/welche/welches, was für einen/eine/ein, wie viel*;
- pronomi relativi (*Relativpronomen*): *der/die/das, welcher/welche/welches, wer/was*.

### *I verbi*

Tra le parole variabili, oltre alle parole declinabili, rientrano anche quelle coniugabili: i verbi. In base a caratteristiche semantiche, si distingue<sup>21</sup> tra verbi pieni, verbi ausiliari, verbi modali, verbi copulativi.

I verbi pieni (*Vollverben* o anche *Hauptverben*) hanno una semantica propria e formano il predicato senza altri verbi.

I verbi ausiliari (*Hilfsverben, Auxiliare*) sono utilizzati per:

- alcuni tempi verbali (*Perfekt* e *Plusquamperfekt* con *sein/haben*; *Futur I* e *Futur II* con *werden*);
- alcuni modi verbali (*Konjunktiv* con *würde*);
- il passivo<sup>22</sup> (*Passiv*).

I verbi modali (*Modalverben*) esprimono possibilità, necessità, permesso, capacità ecc. (*können, dürfen, müssen, sollen, wollen, mögen* + *Infinitiv*). Possono esprimere una modalità deontica (*subjektbezogen, deontisch*), quando definiscono una relazione tra il soggetto e l'oggetto del discorso (obbligo, necessità, permesso ecc.), come in (7):

(7) Was **muss** ich jetzt tun? (deTenTen20)

Possono esprimere anche una modalità epistemica (*sprecherbezogen, epistemisch*), per esempio una probabilità come in (8):

20 La forma del riflessivo è unica. Può essere rafforzata con *selbst* (cfr. Duden 2022: 740).

21 La *Dudengrammatik* (cfr. Duden 2022: 638) distingue due macrocategorie: verbi pieni (*Vollverben*) e verbi che reggono un infinito (*infinitregierende Verben*). I verbi che reggono un infinito si distinguono a loro volta in base alla forma verbale che reggono: se reggono il participio passato (*Partizip II*) → *haben, sein, werden, bekommen* ecc.; se reggono l'infinito senza *zu* → verbi modali e verbi cinestetici (*sehen, hören* ecc.); se reggono l'infinito con il *zu* → *haben, sein, scheinen* ecc.

22 Si distinguono vari tipi di passivo: il passivo formato con l'ausiliare *werden*, detto anche *Vorgangspassiv* (*werden* + *Partizip II*), il passivo di stato *Zustandspassiv* formato con l'ausiliare *sein* (*sein* + *Partizip II*) e il costruito al dativo *Dativpassiv* o *Rezipientenpassiv* formato con gli ausiliari *bekommen, erhalten, kriegen* (*bekommen/erhalten/kriegen* + *Partizip II*). Le forme verbali realizzate con gli ausiliari vengono chiamate *analytische Verbformen*, quelle senza ausiliari *synthetische Verbformen*.

(8) Es **muss** wohl an der Hitze liegen. (deTenTen20)

Simili ai verbi modali sono gli *Halbmodalverben*, che reggono l'infinito preceduto da *zu*, come in (9):

(9) Die Deutschen **scheinen** hier keine Rolle **zu** spielen. (deTenTen20)

Vi sono poi i verbi copulativi (*Kopulaverben*), che indicano uno stato (*sein*) o l'inizio o la continuazione di uno stato (*werden, bleiben*). Insieme ad altri elementi, come gli aggettivi e i sintagmi nominali (*Prädikative*), costituiscono il predicato.

## 2. Parole non variabili:

Le parole non variabili comprendono gli avverbi<sup>23</sup> (*Adverbien*), le preposizioni (*Präpositionen*), le congiunzioni coordinanti (*koordinierende Konjunktionen*), le congiunzioni subordinanti (*subordinierende Konjunktionen*), gli aggiuntori (*Adjunktoren o Vergleichspartikeln*), le particelle (*Partikeln*) e le interiezioni (*Interjektionen*). Per la loro descrizione, dal momento che decade il criterio morfologico, possono essere usati soprattutto criteri distribuzionali e semantici (cfr. Pittner 2016: 94).

### *Gli avverbi*<sup>24</sup>

Gli avverbi sono quelle parole che generalmente precisano il tempo, il luogo, il modo di un evento. Si distinguono in avverbi di tempo, avverbi di luogo, avverbi di modo e causali (gli avverbi di tempo e di luogo sono considerati *Situierungsadverbien* in senso stretto; gli avverbi di modo e causali *Situierungsadverbien* in senso lato).

Gli avverbi di tempo (*temporale Adverbien* o anche *temporale Situierungsadverbien*) precisano il momento temporale di un evento o definiscono una relazione di tipo temporale (momento temporale, durata, frequenza ecc.). Spesso rispondono alle domande *wann? wie lange? wie oft?* Alcuni esempi sono: *heute, eben, jetzt*.

Gli avverbi di luogo (*lokale Adverbien* o anche *lokale Situierungsadverbien*) precisano il luogo di un evento, di un oggetto o definiscono una relazione di tipo spaziale. Si distingue tra avverbi statici di luogo (*Ortsadverbien*), che rispondono alla domanda *wo?* e avverbi di direzione (*Richtungsadverbien*) che contengono spesso *-hin* o *-her* (*wohin? woher?*). Sono una classe molto numerosa: *hier, da, dort, unten, oben, links, rechts* sono alcuni esempi del primo gruppo; *dahin, dorthin, hierin, geradeaus, hinunter, dorher, hierher* sono alcuni esempi del secondo gruppo.

23 Un piccolo gruppo di avverbi è comparabile (cfr. Duden 2022: 791): *oft, wohl* ecc.

24 Oltre alle categorie qui discusse la *Dudengrammatik* (cfr. Duden 2022: 791) aggiunge anche gli avverbi di frase (*Satzadverbien*), che si riferiscono all'intera frase (come, per esempio, *leider, glücklicherweise*) e dal punto di vista semantico esprimono una valutazione da parte di chi parla (cfr. Duden 2022: 798) e gli avverbi con funzione coesiva (*Konjunkionaladverbien*), che uniscono due frasi (come, per esempio, *trotzdem* e *außerdem*).

Gli avverbi di modo (*modale Adverbien*) sono un gruppo molto eterogeneo, anche se al loro interno si possono distinguere avverbi di qualità (*Qualitätsadverbien*) e avverbi di quantità (*Quantitätsadverbien*). Esempi del primo gruppo sono: *so, anders, gern*. Esempi del secondo gruppo sono: *größtenteils; einigermaßen, stundeweise*.

Gli avverbi causali (*kausale Adverbien*) esprimono la causa di un evento o un'azione. Spesso terminano con *-(et)wegen* o *-halber*. Alcuni esempi sono: *meinetwegen, deswegen, gesundheitshalber* (sono causali in senso stretto e rispondono alla domanda *warum?*); *ansonsten, bestenfalls, notfalls* (sono condizionali e rispondono alla domanda *in welchem Fall? unter welcher Bedingung?*); *dazu, dafür* (sono finali e rispondono alla domanda *zu welchem Zweck?*)

Gli avverbi possono costituire da soli un sintagma con la funzione di avverbiale (*Adverbial*) e trovarsi prima di un verbo finito in frasi affermative (cfr. nel *Vorfeld*, campo preposto, si veda il paragrafo 2.1.4). Sono le uniche parole non variabili che possono occupare questa posizione (cfr. Pittner 2016: 94; Duden 2022: 791).

### *Le preposizioni*

Le preposizioni si accompagnano a un gruppo nominale e ne determinano il caso. Senza il gruppo nominale non possono stare da sole nel *Vorfeld* (campo preposto).

Sono una classe numerosa (cfr. Duden 2022: 810); vi sono sia preposizioni semplici (*in, an*) sia complesse (*bezüglich, hinsichtlich*) spesso derivate da aggettivi. In base alla loro posizione rispetto al gruppo nominale si distinguono<sup>25</sup> in preposizioni in senso stretto (*Präpositionen*), come *an, auf*; postposizioni (*Postpositionen*), come *halber, zuliebe*; circumposizioni (*Zirkumpositionen*), come *um...herum, um ...willen*, e ambiposizioni (*Ambipositionen*), utilizzabili sia prima che dopo come *wegen, zufolge*.

Le preposizioni più frequenti sono *mit, von, an, auf* e *zu* (cfr. Duden 2022: 810). Si distinguono in base a criteri semantici in preposizioni di luogo, temporali, causali, modali ecc., ma una preposizione può avere più significati e quindi più usi. Esempi di preposizioni locali sono: *ab, abseits, an, auf, aus*; esempi di preposizioni temporali sono: *ab, an, auf, aus, außerhalb*; esempi di preposizioni causali sono: *angesichts, anlässlich, auf, aus, bei*; esempi di preposizioni modali sono: *abzüglich, anstelle, auf, aus, ausschließlich*.

Le preposizioni reggono normalmente un solo caso: in ordine di frequenza dativo (*ab, aus, außer* ecc.), accusativo (*ausgenommen, betreffend, bis* ecc.), genitivo (*abseits, abzüglich, angesichts* ecc.). Vi sono però preposizioni, come quelle di luogo, che possono reggere sia il dativo, quando indicano uno stato in luogo (10),

25 Tutte questi tipi fanno parte delle *Adpositionen*. Indipendentemente dalla loro posizione rispetto al gruppo nominale si è attestato il termine “preposizione”.

sia l'accusativo quando indicano un movimento (11) (*an, auf, hinter, in, neben, über, unter, vor, zwischen*):

- (10) Sorry, ich habe hier beide Modelle **auf dem Tisch**. (deTenTen20)  
 (11) Setzt das Glas vorsichtig **auf den Tisch**. (deTenTen20)

Alcune preposizioni, come *trotz, wegen* ecc., possono essere usate sia con il dativo sia con il genitivo; l'uso con il genitivo è considerato “più corretto” ed è più diffuso nei testi scritti (12), l'uso con il dativo, invece, più colloquiale e lo si ritrova nella lingua orale o nella CMC (13):

- (12) **Wegen des Lockdowns** werden dort derzeit immer nur wenige Personen zur gleichen Zeit eingelassen. (deTenTen20)  
 (13) Muss man sich schon wieder aufregen, **wegen so einem Blödsinn?** (deTenTen20)

#### *Le congiunzioni coordinanti e le congiunzioni subordinanti*

Le congiunzioni uniscono frasi o parti di frase e non hanno funzione di costituente sintattico (*Satzglied*). Per *Satzglied* si definisce un costituente o sintagma che può occupare da solo il campo preposto (*Vorfeld*) (cfr. Duden 2022: 40).

In base al significato e alla loro funzione si distinguono in congiunzioni additive (*und, sowie, sowohl – als (auch)*), congiunzioni alternative (*oder, entweder – oder, beziehungsweise*), congiunzioni avversative (*aber, doch, sondern*), congiunzioni espletive (*das heißt/ d. h.*).

Si distingue inoltre tra congiunzioni coordinanti (*koordinierende Konjunktionen*) e congiunzioni subordinanti<sup>26</sup> (*subordinierende Konjunktionen*). Le prime coordinano frasi dello stesso livello o porzioni di frase (14), (15):

- (14) Ein Bild ist unterwegs **und** vielen Dank für's Rezept! (deTenTen20)  
 (15) Das gesamte Jahr über gibt es Events **und** Vorträge. (deTenTen20)

Le seconde introducono frasi subordinate (16), (17). Si ricorda che in tedesco, nelle frasi subordinate, il verbo finito occupa l'ultima posizione nella frase:

- (16) Er hofft sehr, **dass** er das Autofahren nochmal schafft. (deTenTen20)  
 (17) **Ob** das reicht, weiß ich nicht. (deTenTen20)

Le congiunzioni subordinanti (*subordinierende Konjunktionen*) sono varie: *das* e *ob*, che introducono frasi dichiarative e interrogative; *wenn, sobald, weil, als, wenn* (avverbiali); *nachdem, während, als, wenn* (temporali); *wenn, falls, im Fall(e)*

<sup>26</sup> In alcune grammatiche si preferisce utilizzare i termini “congiunzioni” (*Konjunktionen*) e “subgiunzioni” (*Subjunktionen*).

(condizionali); *weil, da, wo* (causali); *obwohl, wenngleich* (concessive) e *als, als das* (comparative).

#### *Gli aggiuntori*

Gli aggiuntori (*Adjunktoren* o *Vergleichspartikeln*) non reggono alcun caso e sono la testa di sintagmi aggiuntori (*Adjunktorphrasen*): *als, wie, statt, anstatt, außer* (cfr. Duden 2022: 472). L'uso di *als* è mostrato nell'esempio (18):

(18) Trotzdem ist ein 250er deutlich schneller **als** alle 125er. (deTenTen20)

#### *Le particelle*

Le particelle (*Partikeln*) non occupano una posizione prima del verbo finito (non possono stare nel campo preposto, il *Vorfeld*), non reggono un caso e non hanno la funzione di congiungere (cfr. Pittner 2016: 96-97). Si differenziano in particelle che sono integrate dal punto di vista sintattico e prosodico nella frase e particelle che non lo sono (cfr. Duden 2022: 836). Al primo gruppo appartengono le particelle modali (*Modalpartikeln, Abtönungpartikeln*), le *Fokuspartikeln* e le particelle intensificatrici (*Intensitätspartikeln*). Al secondo le interiezioni (*Interjektionen*) e le particelle responsive (*Antwortpartikeln*).

#### *Le particelle modali*

Le particelle modali (*Modalpartikeln, Abtönungpartikeln*)<sup>27</sup> hanno un significato astratto ed esprimono il punto di vista del parlante sull'enunciato. Sono quasi sempre non accentate<sup>28</sup> e si trovano principalmente nel campo interposto, il *Mittelfeld*, non avendo autonomia sintattica. In particolare, tendono a comparire nella parte più a sinistra del *Mittelfeld* (cfr. Cognola/Moroni 2022: 24), precedendo il focus della frase e seguendo i pronomi non accentati (cfr. Cognola/Moroni 2022: 24; Thurmair 1989: 32-35). Alcuni esempi sono: *aber, auch, bloß, denn, doch*.

Sono collegate al tipo di frase: ci sono particelle che vengono usate preferibilmente in frasi affermative (19), altre in frasi interrogative (20), altre in frasi imperative (21).

(19) Bei Lindner dürfen auch die Sportler „sündigen“, es sind **doch** viele Vitami-

27 Alcuni studiosi (cfr. Zimmermann 2011) e la *Dudengrammatik* (cfr. Duden 2022: 842) usano il termine *Diskurspartikeln*, dal momento che queste unità ancorano una frase nel discorso. Secondo Cognola/Moroni (2022: 22) questa soluzione può dare luogo a fraintendimenti: la funzione delle *Diskurspartikeln* è legata anche alla gestione del turno di parola (Blühdorn 2017); questo aspetto è estraneo però alle *Modalpartikeln*, che hanno anche un comportamento sintattico diverso (Cognola/Moroni 2022: 22).

28 Vi sono eccezioni come per esempio *ja, bloß, nur* in minacce. Dal punto di vista prosodico le particelle modali presentano delle restrizioni. Anche questo aspetto giustifica la loro trattazione come classe funzionale autonoma (Cognola/Moroni 2022: 26).

ne dabei. (deTenTen20)

(20) Äh ... was machst du da **eigentlich**? (deTenTen20)

In base al tipo particella gli atti illocutori possono essere vari; nei seguenti esempi l'ordine espresso dalla frase imperativa (21) diventa un permesso (22) o una minaccia (23).

(21) Bleib hier!

→ Frase imperativa; ordine (cfr. Pittner 2016: 97)

(22) Bleib **ruhig** hier!

→ Frase imperativa; permesso (cfr. Pittner 2016: 97)

(23) Bleib **bloß** hier!

→ Frase imperativa; minaccia (cfr. Pittner 2016: 97)

Le particelle modali si riferiscono a tutta la frase e possono trovarsi anche in frasi secondarie.

Le unità linguistiche usate come particelle possono essere spesso usate anche in altre classi di parola, come aggettivi (*ruhig, eigentlich*), avverbi (*schon, vielleicht*), congiunzioni (*denn, aber*), particelle focalizzanti (*auch, nur*), particelle responsive *Antwortpartikeln* (*ja, doch*).

#### *Le Fokuspartikeln*

Le *Fokuspartikeln*<sup>29</sup> (particelle focalizzanti) sono particelle che hanno un significato quantitativo (*quantifizierend*) e fanno spesso riferimento a una scala di valori (cfr. Duden 2022: 837). Formano un costituente sintattico con l'elemento portatore dell'accento di frase.

Le particelle focalizzanti sono relativamente libere nella frase, come mostrano gli esempi con *nur* (24), (25), (26):

(24) Mobbing ist nicht **nur** auf dem Schulhof ein Thema. (deTenTen20)

(25) **Nur** Mobbing ist nicht auf dem Schulhof ein Thema. (deTenTen20)

(26) Mobbing ist nicht auf dem Schulhof **nur** ein Thema. (deTenTen20)

Nella frase (24) *nur* si riferisce a *auf dem Schulhof*, nella (25) a *Mobbing*, nella (26) a *ein Thema*. Il costituente, che di norma segue la particella<sup>30</sup>, è accentato e costituisce l'informazione più importante. Viene dunque focalizzato (da qui il nome di *Fokuspartikeln*).

Le principali *Fokuspartikeln* sono: *auch, nur, sogar*. Alcune di esse sono restrittive (*nur, bloß, allein*), altre additive (*auch, ebenfalls, ebenso*). Altre, invece, non rientrano in nessuna delle due categorie: *ausgerechnet, eben, gerade*. Anche *nicht* è considerata una *Fokuspartikel*.

29 In alcune grammatiche vengono chiamate *Gradpartikeln* (cfr. Duden 2022: 837).

30 In rari casi il costituente precede la particella o è separato dalla particella da una o più parole.

La relazione con il costituente è varia (27), (28), (29):

(27) **Auch** Oma wird kommen. (deTenTen20)

(28) **Nur** Oma wird kommen.

(29) **Sogar** Oma wird kommen.

In (27) vi è almeno una alternativa (oltre alla nonna verranno anche altre persone); in (28) le altre alternative vengono escluse (verrà solo la nonna); in (29) viene espressa una valutazione, ossia che inaspettatamente verrà anche la nonna.

#### *Le particelle intensificatrici*

Le particelle intensificatrici (*Intensitätspartikeln*, *Steigerungspartikeln*, *Intensivpartikeln*, *Intensivierungspartikeln*) indicano il grado della caratteristica espressa dall'elemento a cui si accompagnano (aggettivo, verbo, avverbio) (30), (31);

(30) Sieht **ziemlich** spektakulär aus. (deTenTen20)

(31) Mama, ich liebe dich so **sehr!** (deTenTen20)

Alcuni esempi sono: *sehr*, *zu*, *allzu*. Dal punto di vista semantico si distingue tra particelle che rafforzano e particelle che attenuano<sup>31</sup>.

Senza l'elemento a cui fanno riferimento non possono precedere il verbo finito; se si riferiscono alla frase, invece, in alcuni casi possono essere spostate nel campo preposto, il *Vorfeld*. Alcune particelle intensificatrici possono essere usate solo con certi tipi di sintagmi, per esempio *sehr* non si unisce quasi mai con sintagmi preposizionali usati in funzione di avverbiale (cfr. Duden 2022: 841). Le particelle intensificatrici possono essere portatrici di accento.

#### *Le particelle responsive*

Le particelle responsive (*Antwortpartikeln*) si usano in risposta a domande *Ja/Nein* (*Entscheidungsfragen*) (32), (33)<sup>32</sup>.

(32) Gibt es Studenten-Abos? **Ja**, die gibt es. (deTenTen20)

(33) Sind Rückerstattungen möglich? **Nein**, die Eintrittskarte kann nach dem Kauf nicht zurückgegeben werden. (deTenTen20)

31 Breindl (2009) distingue invece le particelle che hanno un riferimento all'interno di una scala di valori (*Skalenbereich*) come *überaus*, *sehr*, *ein bisschen* e particelle che sono al limite della scala, come *fast*, *beinahe*, *vollends*.

32 In risposta possono essere usati anche avverbi di frase (*Satzadverbien*). A differenza delle particelle responsive, gli avverbi di frase (*Satzadverbien*), possono occupare il campo preposto, il *Vorfeld*, di una frase affermativa.

### Le interiezioni

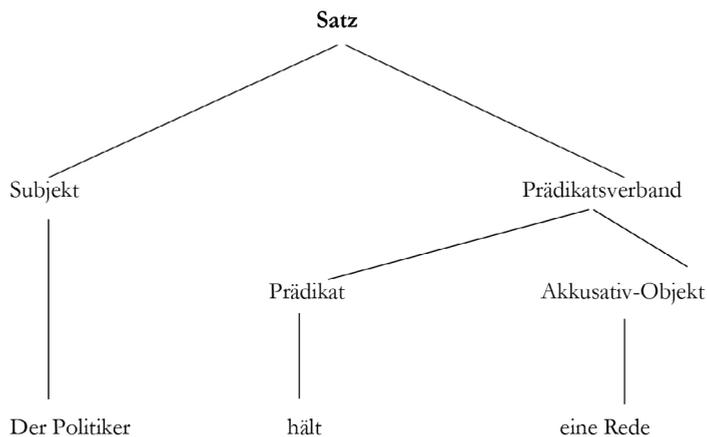
Le interiezioni (*Interjektionen*), oltre ad avere funzione espressiva, hanno funzione discorsiva, e vengono dunque definite *Gesprächspartikeln*. Hanno un proprio contorno intonativo e sono isolate da un punto di vista sintattico. Il loro status di parola è discusso, dato che non sempre hanno la struttura sillabica tipica del tedesco. Esempi sono: *mmb, na ja, brr, gell, aua, igitt, Mensch, Mist*.

### 2.1.2 Sintagmi

Per formare una frase, le parole (le unità elementari) si combinano, portando con sé le caratteristiche della classe a cui appartengono, in gruppi di parole (*Wortgruppen*), detti anche in senso generico costituenti (*Konstituenten*).

Una frase può essere analizzata e descritta in base a diversi criteri e modalità<sup>33</sup> (cfr. Busch/Stenschke 2018: 125-126). In base al sistema tradizionale, che si concentra su strutture e gerarchie, la frase *Der Politiker hält eine Rede* può essere suddivisa nel seguente modo (34):

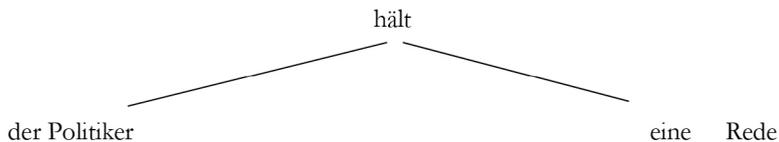
(34) Der Politiker hält eine Rede.



In base al modello della grammatica della dipendenza e valenza, che considera il verbo finito la struttura centrale della frase, la frase (35) può essere analizzata nel seguente modo:

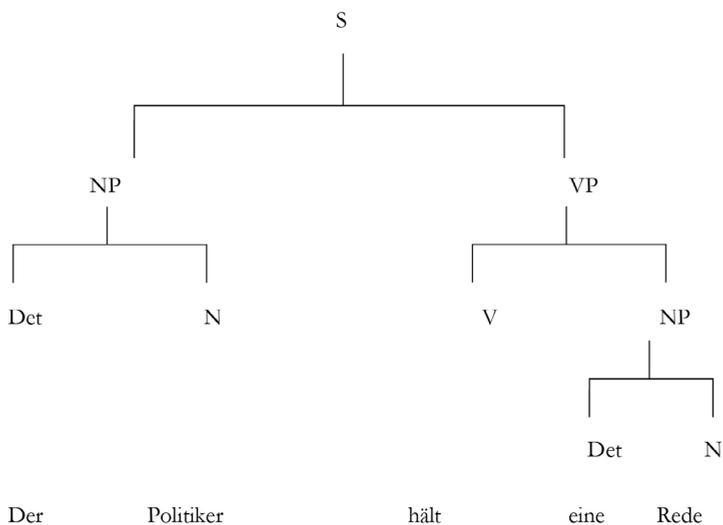
<sup>33</sup> Gaeta (2017: 177) afferma che sostanzialmente esistono due impostazioni: i modelli universalisti, che riducono gli schemi osservati a una struttura profonda (*Tiefenstruktur*), da cui derivano tramite regole di movimento le varietà della struttura superficiale (*Oberflächenstruktur*) e i modelli più orientati alla descrizione concreta degli schemi sintattici.

(35) Der Politiker hält eine Rede.



Dal punto di vista di altre grammatiche, che prendono in esame le strutture dei sintagmi (come, per esempio, anche la grammatica generativa), la frase può essere descritta mediante un diagramma ad albero (36):

(36) Der Politiker hält eine Rede.



Legenda

Det = *Determinativ* (articolo)

N = *Nomen* (nome, sostantivo)

S = *Satz* (frase)

NP = *Nominalphrase* (sintagma nominale)

VP = *Verbalphrase* (sintagma verbale)

Un altro tipo di rappresentazione alternativa è la notazione gerarchizzata dei sintagmi mediante parentesi quadra (37):

(37) Der Politiker hält eine Rede.

[[Der Politiker]<sub>NP</sub> [[hält]<sub>V</sub> [eine Rede]<sub>NP</sub>]<sub>VP</sub>]<sub>S</sub>

Ogni gruppo di parola può essere classificato in base alle sue caratteristiche grammaticali in tipi o categorie chiamati sintagmi (*Phrasen*). La classificazione avviene in base al loro elemento principale e obbligatorio, ossia la testa (I) (*Kern, Kopf*).

In base alla testa si distinguono i seguenti tipi di sintagmi (Tabella 2.2):

Tipo di sintagma	Testa
sintagma nominale ( <i>Nominalphrase, NP</i> )	nome, pronomi
sintagma preposizionale ( <i>Präpositionalphrase, PP</i> )	preposizione
sintagma verbale ( <i>Verbalphrase, VP</i> )	verbo
sintagma aggettivale ( <i>Adjektivphrase, AdjP</i> )	aggettivo
sintagma avverbiale ( <i>Adverbphrase, AdvP</i> )	avverbio
sintagma aggiuntore ( <i>Adjunktorphrase, AdjkP</i> )	aggiuntore

**Tabella 2.2:** Tipo di sintagma e indicazione della relativa testa<sup>34</sup>

Nel **sintagma nominale** (*Nominalphrase, NP*) la testa è un sostantivo o un pronome, ed è espandibile con altri elementi, come per esempio articoli, aggettivi, altri sintagmi nominali al genitivo, sintagmi preposizionali e frasi. Esempi di sintagma nominale sono (la testa è sottolineata) in (38):

(38)

Schildkröte; die Schildkröte; eine grüne Schildkröte; die grüne Schildkröte; sie; die Schildkröte meines Bruders; die Schildkröte von meinem Bruder; die Schildkröte, die wir gestern im Garten gefunden haben.

Tutti questi sintagmi sono interscambiabili in base al criterio distribuzionale. Gli elementi usati per espandere il sintagma nominale sono definiti attributi; si distingue tra:

- attributo aggettivale (*Adjektiv-Attribut*), come in *eine grüne Schildkröte*;
- attributo al genitivo (*Genitiv-Attribut*), come in *die Schildkröte meines Bruders*;
- attributo preposizionale (*Präpositional-Attribut*), come in *die Schildkröte von meinem Bruder*;
- frase attributiva (*Attributsatz*), come in *die Schildkröte, die wir gestern im Garten gefunden haben*;
- apposizione (*Apposition*), come in *Onkel Dirk*.

<sup>34</sup> Cfr. anche altre classificazioni, come in Bianco/Mollica (2024) e in *Grammis* (<https://grammis.ids-mannheim.de/progr/mm/5213>). In queste due classificazioni il sintagma nominale viene differenziato dal sintagma pronominale e scompare la denominazione di sintagma verbale, preferendo l'uso del termine "complesso verbale" (*Verbalkomplex*).

Gli elementi presenti in un sintagma presentano una relazione di tipo grammaticale (*Kongruenz*)<sup>35</sup>. In un sintagma nominale il genere è determinato dal sostantivo; le classi di parola modificabili (si vedano il paragrafo 1.1.1 e il paragrafo 2.1.1 della PARTE II) devono essere in accordo rispetto a caso e numero.

Nel **sintagma preposizionale** (*Präpositionalphrase, PP*) la testa è una preposizione; questa regge di norma un sintagma nominale, determinandone il caso (*Rektion*)<sup>36</sup>. Sono possibili tutti i tipi di sintagma nominale presentati precedentemente. Esempi di sintagma preposizionale sono (la testa è sottolineata) (39):

(39)  
auf dem Tisch; auf dem alten Tisch; auf dem Tisch, der vom Flohmarkt kommt;  
auf ihm.

Nel sintagma preposizionale possono trovarsi anche avverbi come in *ab* [*soufort*]<sub>Advp</sub> o altri sintagmi preposizionali come in *bis* [*auf den heutigen Tag*]<sub>PP</sub>

Nei **sintagmi verbali** (*Verbalphrase, VP*) la testa è un verbo semanticamente pieno (*Vollverb*) o un verbo copulativo (*Kopulaverb*). I verbi semanticamente pieni svolgono un ruolo centrale nella frase «da sie bestimmte Leerstellen haben, die durch Ergänzungen gefüllt werden» (Pittner 2016: 101). Questa caratteristica dei verbi viene chiamata valenza (*Valenz*) (cfr. anche Pittner 2016: 101).

Esempi di sintagma verbale sono (la testa è sottolineata) (40):

(40)  
 gerne Cola trinken; im See baden; irritiert sein.

Nei **sintagmi aggettivali** (*Adjektivphrase, AdjP*) la testa è un aggettivo. Gli aggettivi possono essere accompagnati anche da intensificatori e, come i verbi, possono determinare una serie di posizioni da riempire (*Akkusativergänzung, Dativergänzung, Genitivergänzung, Präpositionalergänzung*). Esempi di sintagma aggettivale sono (la testa è sottolineata) (41):

(41)  
 sehr groß; auf sie sauer; ihrem Bruder ähnlich; hundert Euro wert; des Deutschen mächtig

Nei **sintagmi avverbiali** (*Adverbphrase, AdvP*) la testa è un avverbio. Gli avverbi possono essere accompagnati anche da intensificatori, da sintagmi preposizionali e da altri avverbi. Esempi di sintagma avverbiale sono (la testa è sottolineata) (42):

35 Per congruenza si intende «eine regelhafte Übereinstimmung zwischen Elementen in bestimmten grammatischen Merkmalen» (Pittner 2016: 101).

36 Quando un elemento determina le caratteristiche grammaticali di un altro elemento dipendente da lui (retto) si parla di *Rektion*.

(42)  
unten am Bach; dort im Ofen

Nei **sintagmi aggiuntori** (*Adjunktorphrase*, *AdjkP*) la testa è un aggiunto-re (*Adjunktor* o *Vergleichspartikel*).

(43) als Gruppe; wie schrecklich

Per determinare le categorie di sintagmi si applicano criteri distribuzionali (cfr. Pittner 2016: 100): appartengono al tipo sintagma nominale tutti i sintagmi che occupano la medesima posizione all'interno della frase (44):

(44)  
 die Katze  
 die rote Katze  
 die Katze auf dem Dach gehört Sara  
 die Katze, die miaut

La testa di un sintagma è l'elemento centrale del sintagma stesso, che determina anche il suo nome. Il resto del sintagma si definisce "complemento"<sup>37</sup>, ed è l'elemento che precisa la testa. Il complemento è sintatticamente dipendente dalla testa.

Nel tedesco contemporaneo l'ordine di testa (T) e complemento (C) non è sempre lo stesso in tutti i sintagmi<sup>38</sup>.

Nel **sintagma nominale** prevale l'ordine TC (la testa è sottolineata) (45):

(45)  
das Auto des Vaters; die Geschäfte in Rom; der Wunsch nach Frieden; die Tank-  
stelle dort; die Frau, die gestern Schokolade gekauft hat

Fanno eccezione pochi casi, come *Alexanders Rucksack*, in cui l'ordine è CT.

Nel **sintagma aggettivale**, se l'aggettivo è usato in funzione attributiva, l'ordine è CT: *die [sehr schöne] Tasche*. Il complemento può essere un sintagma avverbiale o un sintagma preposizionale. Se l'aggettivo è usato in funzione predicativa, sono possibili entrambi gli ordini (la testa è sottolineata): CT come in *sie ist [an Musik interessiert]* e TC come in *sie ist [interessiert an Musik]*.

Nel **sintagma avverbiale** l'ordine, invece, è sempre CT (la testa è sottolineata) (46):

(46)  
 sie schwimmt [sehr schnell]; er schläft [sehr oft] im Auto ein.

37 Il termine "complemento" viene utilizzato in linguistica con diverse accezioni; in questo contesto è inteso come elemento che precisa la testa di un sintagma.

38 Alcuni fenomeni analizzati da una prospettiva diacronica hanno messo in evidenza che l'ordine principale era originariamente TC (cfr. Di Meola 2014: 104-105).

Nel **sintagma preposizionale**<sup>39</sup> l'ordine più diffuso è TC: *vor dem Haus*; *im Haus*; *entlang des Flusses*. Vi sono però preposizioni che ammettono entrambi gli ordini: TC come in *des Angeklagten wegen* o *dem Kunden gegenüber*. Per altri sintagmi preposizionali, invece, è possibile solo l'ordine CT: (*der Umwelt zuliebe*; *der Vollständigkeit halber*). Sono di fatto postposizioni.

Il **sintagma verbale** mostra l'ordine TC nella frase principale e CT nella frase secondaria introdotta da congiunzione subordinante (47):

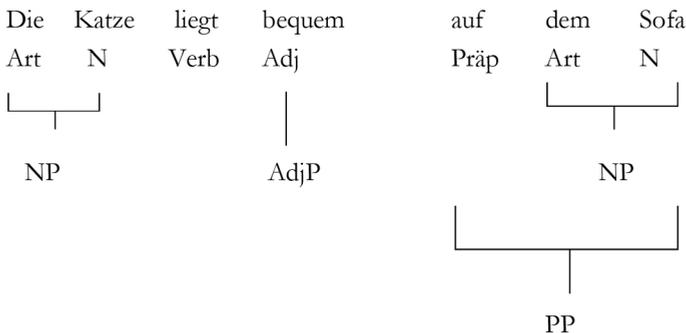
- (47)  
 er kommt nicht;  
 sie kann Klavier spielen;  
 sie kommt nicht, weil sie Klavier spielen muss.

Il verbo coniugato (anche l'ausiliare quindi) è la testa.

Nel **sintagma aggiuntore** l'ordine è TC: *als Gruppe*; *wie schrecklich*.

Riassumendo: le parole o gruppi di parole che formano una frase sono definite genericamente costituenti. Questi possono essere classificati in base a caratteristiche grammaticali in sintagmi (*Phrasen*). Qui di seguito la suddivisione di una frase in costituenti per mezzo di un diagramma ad albero (48):

- (48) Die Katze liegt bequem auf dem Sofa.



Legenda

Art = *Artikel* (articolo)

N = *Nomen* (nome, sostantivo)

Adj = *Adjektiv* (aggettivo)

Präp = *Präposition* (preposizione)

NP = *Nominalphrase* (sintagma nominale)

PP = *Präpositionalphrase* (sintagma preposizionale)

AdjP = *Adjektivphrase* (sintagma aggettivale)

39 Secondo Gaeta (2017: 188) il nome corretto sarebbe sintagma adposizionale (*Adpositionalphrase*) per l'ordine degli elementi possibile: CT e TC.

Per identificare i costituenti sintattici di una frase possono essere svolti una serie di test sintattici (*Konstituententests*)<sup>40</sup>:

1. il test di domanda (*Fragetest*), in base al quale tutto ciò che può essere “chiesto” per mezzo di una domanda probabilmente è un costituente sintattico;
2. il test di sostituzione (*Ersatzprobe, Substitutionstest*), chiamato anche test di pronominalizzazione (*Pronominalisierungstest, Pronominalisierungsprobe*), secondo cui tutto ciò che può essere sostituito mantenendo la grammaticalità della frase probabilmente è un costituente sintattico. Nel caso del test di pronominalizzazione la sostituzione avviene per mezzo di un pronome. La sostituzione può avvenire anche mediante pronomi interrogativi (in questo caso è anche un test di permutazione) o anche mediante espressioni sinonimiche. Esiste anche il test di cancellazione (*Weglassprobe*). Non tutti gli elementi possono essere sempre cancellati, come per esempio i costituenti che svolgono la funzione di soggetto, alcuni costituenti con funzione di oggetto diretto (*sie beantwortet die Fragen* = non può essere cancellato *die Fragen*), alcuni avverbiali (*er wohnt seit 4 Monaten im selben Haus* = non può essere cancellato *im selben Haus*) ecc.;
3. il test di permutazione (spostamento, movimento) (*Verschiebetest, Verschiebeprobe, Umstellprobe*), per il quale tutto ciò che può essere spostato lasciando inalterata la grammaticalità della frase probabilmente è un costituente sintattico;
4. il test di topicalizzazione (*Vorfeldtest*), in base a cui tutto ciò che può essere spostato nel *Vorfeld* è un costituente sintattico<sup>41</sup>.

Qui di seguito una applicazione dei test sopraelencati alla frase *Die Katze liegt bequem auf dem Sofa*:

1. Secondo il test di domanda (*Fragetest*) i costituenti sintattici della frase *Die Katze liegt bequem auf dem Sofa* sono:
  - *die Katze*: risponde alla domanda con il pronome interrogativo *wer?*
  - *bequem*: risponde alla domanda con il pronome interrogativo *wie?*
  - *auf dem Sofa*: risponde alla domanda con il pronome interrogativo *wo?*
  - *dem Sofa*: risponde alla domanda con il pronome interrogativo *wem?*
2. Secondo il test di pronominalizzazione (*Pronominalisierungstest, Pronominalisierungsprobe*), i costituenti sintattici della frase *Die Katze liegt bequem auf dem Sofa* sono:

40 Busch/Stenschke (2018: 127) introducono il *VEA-Prinzip* (*V* = *Verschiebung*; *E* = *Ersetzung*; *A* = *Anfangsstellung*) per identificare i costituenti sintattici. Di Meola (2014: 93-94) e Gaeta (2017: 177) inseriscono anche il test di coordinazione (*Koordinationsprobe*): se due elementi si possono coordinare probabilmente si tratta di costituenti.

41 Essendo il test di topicalizzazione, di fatto, uno spostamento, esso potrebbe essere considerato un tipo specifico di permutazione.

- *die Katze*: può essere sostituito con il pronome *sie*;
  - *bequem*: può essere sostituito con il pronome *so*;
  - *auf dem Sofa*: può essere sostituito con il pronome *darauf*;
  - *dem Sofa*: può essere sostituito con il pronome *ihm*.
3. Secondo il test di permutazione (*Verschiebetest*, *Verschiebeprobe*, *Umstellprobe*), i costituenti sintattici della frase *Die Katze liegt bequem auf dem Sofa* sono:
- *die Katze*: **Die Katze** liegt bequem auf dem Sofa → Bequem liegt die **Katze** auf dem Sofa;
  - *bequem*: Die Katze liegt **bequem** auf dem Sofa → **Bequem** liegt die Katze auf dem Sofa;
  - *auf dem Sofa*: Die Katze liegt bequem **auf dem Sofa** → **Auf dem Sofa** liegt die Katze bequem.

*Dem Sofa* non è un costituente sintattico dal momento che non può essere spostato da solo.

4. Secondo il test di topicalizzazione (*Vorfeldtest*), i costituenti sintattici della frase *Die Katze liegt bequem auf dem Sofa* sono:
- *die Katze* → **Die Katze** liegt bequem auf dem Sofa;
  - *bequem* → **Bequem** liegt die Katze auf dem Sofa;
  - *auf dem Sofa* → **Auf dem Sofa** liegt die Katze bequem.

*Dem Sofa* da solo non è un costituente sintattico dato che non può essere spostato nel *Vorfeld*.

Tutti i costituenti che soddisfano questi test vengono denominati costituenti sintattici (*Satzglieder*). I costituenti in senso generico (*Konstituenten*) e i costituenti sintattici (*Satzglieder*) non sono la stessa cosa: come è stato dimostrato in 3. e in 4. *dem Sofa* è un costituente in senso generico, ma non un costituente sintattico (*Satzglied*) perché non supera il test.

I costituenti sintattici non possono però essere identificati in maniera univoca in tutte le frasi, essendo soggetti a più classificazioni, come nella seguente frase (46):

(46) Lara beobachtet die Frau mit dem Fernglas.

In questa frase i costituenti sintattici possono essere:

- *Lara*; *die Frau*; *mit dem Fernglas* dove il binocolo (*Fernglas*) è lo strumento con cui Lara vede la donna;
- *Lara*; *die Frau mit dem Fernglas* dove è la donna ad avere lo strumento in mano.

In questi casi si parla di ambiguità lessicale e strutturale.

Inoltre, esistono casi in cui un test può essere superato e un altro no; quindi, solo un numero significativo di test può determinare se un costituente è un costituente sintattico (cfr. Di Meola 2014: 96).

### 2.1.3 Funzioni sintattiche nella frase semplice

Gli elementi che costituiscono una frase sono in relazione tra di loro e svolgono una precisa funzione<sup>42</sup>. Si prenda ad esempio la frase (47):

(47) Ein Huhn trinkt Wasser. (de'TenTen20)

Vi sono due sintagmi nominali: *ein Huhn* e *Wasser*. I due hanno una relazione diversa con il verbo *trinken*, un verbo pieno (*Vollverb*):

- *ein Huhn* svolge la funzione di soggetto;
- *Wasser* svolge la funzione di oggetto.

Come già accennato nel paragrafo 2.1.2, i costituenti che superano tutti i test sintattici vengono chiamati costituenti sintattici (*Satzglieder*). I costituenti sintattici possono svolgere la funzione di soggetto, oggetto, avverbiale e predicativo<sup>43</sup>. Non vi è però un rapporto 1:1 tra forma (categoria sintattica) e funzione: un sintagma può cioè avere più funzioni. Il sintagma nominale, per esempio, può svolgere la funzione di soggetto, oggetto, avverbiale, predicativo e attributo. Dall'altro lato, una funzione può essere assolta da più sintagmi, come nel caso della funzione di avverbiale, che può essere svolta dal sintagma avverbiale, dal sintagma aggettivale, dal sintagma nominale, dal sintagma preposizionale e da una frase (cfr. Pittner 2016: 106). Per identificare le funzioni dei *Satzglieder* sono molto utili alcuni dei test menzionati sopra, il test di domanda (*Fragetest*) e il test di pronominalizzazione (*Pronominalisierungstest*, *Pronominalisierungsprobe*).

È importante sottolineare nuovamente che la funzione dei costituenti sintattici dipende dal verbo, che, come già anticipato, ha un ruolo fondamentale all'interno della frase, determinando l'organizzazione e il tipo dei costituenti sintattici necessari. Questa caratteristica del verbo, chiamata valenza, risale a Tesnière (1959), che ha adottato in linguistica alcuni concetti della chimica, per esempio, la caratteristica di alcuni elementi chimici di attrarne altri. Per illustrare meglio la sua teoria, Tesnière usa anche la metafora del teatro: per rappresentare un pezzo teatrale vanno in scena gli attanti (*Aktanten*) con una serie di scenografie (*Kulissen*) che situano gli avvenimenti da un punto di vista spaziale, temporale. Questi vengono chiamati anche circostanziali (*circostants*)<sup>44</sup>.

42 In questo paragrafo si focalizza solo la funzione sintattica. I sintagmi svolgono però anche una funzione semantica.

43 Per ogni funzione possono essere elencati elementi prototipici (cfr. Busch/Stenschke 2018: 133). Dürscheid (2012: 31-39) non considera il *Prädikativum* un costituente sintattico, ma parte del predicato.

44 Nel corso del tempo sono state pubblicate anche delle grammatiche e dizionari che si basano sul modello della valenza. Per le grammatiche cfr. Engel (2009) e per nuovi sviluppi del modello con adattamenti al tedesco contemporaneo in chiave contrastiva cfr. Bianco/Mollica (2024). Per i dizionari cfr. *E-valbu. Valenzwörterbuch deutscher Verben*, <https://grammis.ids-manheim.de/verbvalenz> [12.12.23].

Dalla valenza del verbo si determinano il numero e il tipo degli *Aktanten*, detti anche *Ergänzungen* o *Komplemente* (in italiano, complementi di verbo). Soggetto e oggetto sono solitamente *Ergänzungen*.

Con i verbi che hanno valenza 1, come *niesen* e *busten*, è obbligatoria una sola *Ergänzung* (48); con i verbi con valenza 2, come *kaufen*, due *Ergänzungen* (49); con i verbi con valenza 3, come *schenken* (50), tre *Ergänzungen*, e così via:

(48) Das Kind **niest** und **hustet** (deTenTen20)

(49) Wir **kaufen** PKW's. (deTenTen20)

(50) Ich **schenke** dir ein E-Book. (deTenTen20)

Nell'esempio (48) *das Kind* è il soggetto del verbo *niest* e del verbo *hustet*, è il complemento di verbo (*Ergänzung*) obbligatorio. Nell'esempio (49) *wir* è il soggetto del verbo *kaufen*, mentre *PKW's* è l'oggetto diretto. Entrambi sono complementi di verbo (*Ergänzungen*) obbligatori. Nell'esempio (50) *ich* è il soggetto del verbo *schenke*, mentre *dir* è l'oggetto indiretto e *ein E-Book* l'oggetto diretto. *Ich*, *dir* e *ein E-Book* sono complementi di verbo (*Ergänzungen*) obbligatori. Le frasi costituite solo dal verbo (portatore di valenza) e i suoi attanti, ossia complementi di verbo, sono chiamate frasi nucleari, essendo costituite dal nucleo dell'enunciato (cfr. Bianco/Mollica 2024).

I circostanziali (*Angaben*) aggiungono informazioni, di tipo locale, temporale, sulla modalità ecc.; non sono specifici di una sottoclasse verbale e possono quindi ricorrere con quasi tutti i verbi, non sono dipendenti dal verbo e spesso possono essere anche tralasciati. Come indicato sopra i sintagmi avverbiali possono trovarsi sia in funzione di complemento di verbo (cioè attanziale), sia di circostanziale.

La distinzione tra *Ergänzungen* (complementi di verbo) e *Angaben* (circostanziali) non è sempre semplice, dato che talvolta anche le *Ergänzungen* possono essere omesse (Dürscheid 2012). Si parla, pertanto, di un continuum tra i rappresentanti più prototipici delle due classi (cfr. Bianco/Mollica 2024).

### *Soggetto*

Il soggetto risponde alle domande *wer? was?*

Di norma è un sintagma nominale al nominativo<sup>45</sup> e concorda per numero e persona con il verbo coniugato (esempi 51 e 52):

(51) **Miriam** (3. Person Singular) schwimmt. (3. Person Singular)

(52) **Die Kinder** (3. Person Plural) schwimmen. (3. Person Plural)

45 Non tutti i sintagmi al nominativo sono però soggetti, come si evince dalla seguente frase: *Albert Einstein ist ein Physiker*. *Albert Einstein* è il soggetto della frase, mentre *ein Physiker* no.

Nella frase (53) *untersuchten* è il verbo, mentre *die Experten* è il soggetto (risponde alla domanda *wer?*):

(53) **Die Experten** untersuchten 27 Sorten Balsamicoessig. (deTenTen20)

Quasi tutti i verbi del tedesco necessitano del soggetto; sono pochissimi i verbi che possono essere usati senza soggetto (*frieren, grauen, dursten* ecc.). In (54) un esempio con *frieren*.

(54) Mich **friert**, mich **friert!** Ich möchte zu Hause sein! (deTenTen20)

Con l'infinito e il participio il soggetto è raro; con l'imperativo non è usuale, ma può essere presente.

### Oggetto

L'oggetto è determinato dal verbo (dalla sua valenza), che fissa quanti e quali oggetti debbano essere presenti nella frase. Dal verbo dipende anche il caso o la preposizione.

Si distingue tra:

- oggetto diretto (*Akkusativobjekt, Akkusativergänzung*), che compare all'accusativo. È quello più frequente;
- oggetto indiretto, che può essere al dativo (*Dativobjekt, Dativergänzung*) o più raramente al genitivo (*Genitivobjekt, Genitivergänzung*);
- oggetto preposizionale<sup>46</sup> (*Präpositionalobjekt, Präpositionalergänzung*), che consiste in una preposizione fissata idiomaticamente dal verbo e da un sintagma nominale.

Anche una frase completiva oggettiva può svolgere la funzione di oggetto (cfr. Gaeta 2017: 201), come si evince dall'esempio (55):

(55) Wir hoffen, **dass für Ihren Geschmack ein Schnäppchen dabei ist.**  
(deTenTen20)

### Oggetto diretto

L'oggetto diretto (*Akkusativobjekt, Akkusativergänzung*) compare all'accusativo e risponde alle domande *wen? was?* (56):

(56) Die Experten untersuchten **27 Sorten Balsamicoessig.** (deTenTen20)

Nella frase d'esempio il verbo è *untersuchten*, il soggetto *die Experten* (risponde alla domanda *wer?*) e l'oggetto diretto *27 Sorten Balsamicoessig* (risponde alla domanda *was?*).

<sup>46</sup> Nel test di domanda viene inserita la preposizione.

Se la frase viene trasformata al passivo, l'oggetto diretto assume il ruolo di soggetto (57).

(57) **27 Sorten Balsamicoessig** wurden von den Experten untersucht.

#### *Oggetto indiretto*

L'oggetto indiretto può essere al dativo (*Dativobjekt*) o più raramente al genitivo (*Genitivobjekt*). Risponde alle domande *wem? was?* per il dativo (soprattutto *wem*) e *wessen?* per il genitivo, e indica nella maggior parte dei casi una persona. Se la frase viene trasformata al passivo (*Vorgangspassiv* e *Zustandspassiv*) il suo ruolo non si modifica. Solo nel caso del *Rezipientenpassiv* (con *kriegen* o *bekommen*) assume la funzione di soggetto (58):

(58) Sie schenkte mir Schwimmflügel. (deTenTen20) → **Ich** bekam von ihr Schwimmflügel geschenkt.

L'oggetto indiretto al genitivo è raro; nel tedesco contemporaneo i verbi che reggono l'oggetto indiretto al genitivo sono poco frequenti (59) e in alcuni casi si preferisce usare altri oggetti, come quello preposizionale (60).

(59) erinnern sich: **ihrer**

(60) erinnern sich: **an sie**

Se la frase viene trasformata al passivo, il suo ruolo non si modifica.

Moltissimi verbi in tedesco reggono sia un oggetto diretto, sia un oggetto indiretto, come il verbo *geben* nell'esempio (61): l'oggetto diretto è *die Grundzüge ihrer jetzigen Verfassung*, mentre l'oggetto indiretto è *der Stadt*.

(61) König Johann **gab** der Stadt 1210 die Grundzüge ihrer jetzigen Verfassung. (deTenTen20)

Meno numerosi sono invece i verbi che reggono esclusivamente un oggetto al dativo, come il verbo *danken* nell'esempio (62): l'oggetto al dativo è *ihm*.

(62) Ich **danke** ihm jeden Tag [aufs Neue]. (deTenTen20)

Rari sono invece i verbi che reggono l'oggetto al genitivo, come nell'esempio (63), in cui l'oggetto al genitivo è *ihrer*.

(63) Und ich **schäme** ihrer nicht. (deTenTen20)

Dall'oggetto al dativo deve essere distinto il dativo libero (cfr. Di Meola 2014: 116). L'oggetto al dativo è dipendente dal verbo; riprendendo l'esempio (61), se *der Stadt* non venisse espresso, la frase non sarebbe completa.

Il dativo libero invece non dipende dal verbo, come *ibr* nell'esempio (64):

(64) Sabine öffnet **ibr** die Tür. (deTenTen20)

Da un punto di vista semantico, il dativo libero si suddivide in *dativus commodi* come nell'esempio (65) in *dativus incommodi* (66), in *dativus possessivus* (67), in *dativus indicantis* (68), in *dativus ethicus* (69).

Nell'esempio (65) il *dativus commodi* indica l'entità che risulta avvantaggiata dall'azione e può essere parafrasato con *zum Vorteil von /für*:

(65) Sabine öffnet **ibr** die Tür. (deTenTen20)  
= Sabine öffnet zu ihrem Vorteil (für sie) die Tür.

Nell'esempio (66) il *dativus incommodi* indica l'entità che risulta svantaggiata dall'azione verbale e può essere parafrasato con *zum Nachteil von*:

(66) Das Essen brannte **ibr** an. (deTenTen20)  
= Das Essen brannte zu ihrem Nachteil an.

Il *dativus possessivus* indica un'entità che è in possesso di un'altra o un'entità che fa parte di un'altra (67); può essere sostituito da un pronome possessivo:

(67) Er wäscht **ibr** jetzt die Farbe vom Gesicht. (deTenTen20)  
= Er wäscht jetzt ihre Farbe vom Gesicht.

Nell'esempio (68) il *dativus indicantis* indica un'opinione o un punto di vista personale; può essere sostituito con *meiner Ansicht nach, in meinen Augen*.

(68) War **mir** einfach zu schwierig. (deTenTen20)

Il *dativus ethicus* indica una persona coinvolta emotivamente nell'azione ed è parafrasabile con *mir zyliebe*. Essendo sottinteso un obbligo morale, viene definito dativo etico (esempio 69).

(69) Jetzt fall **mir** bloß nicht um. (deTenTen20)

#### *Oggetto preposizionale*

L'oggetto preposizionale (*Präpositionalobjekt*) è formato da una preposizione, fissata idiomaticamente dal verbo e ormai semanticamente "ridotta", e da un sintagma nominale (70).

(70) Ich warte **auf dich!** (deTenTen20)

Per fare il test di domanda si usa la preposizione. Nell'esempio (70): *auf wen? worauf?*

Il verbo non regge un caso, ma una determinata preposizione.

L'uso di *auf in auf meinem Hof* nell'esempio (71) si distingue da quello di *auf in auf dich*: in questo caso è avverbiale.

(71) Ich warte **auf meinem Hof auf dich**. (deTenTen20)

A livello semantico *auf* indica una dimensione spaziale e potrebbe essere sostituito da altre preposizioni di luogo (*neben, hinter* ecc.). Nel caso di *auf dich* la preposizione non può essere invece sostituita da un'altra preposizione. Sono rarissimi i casi in cui un oggetto proposizionale può essere usato con diverse preposizioni (72) e (73):

(72) Kommt vorbei – wir **freuen** uns **auf euch!** (deTenTen20)

(73) Wir **freuen** uns **über jeden neuen Eintrag!** (deTenTen20)

#### *Complemento avverbiale*

Il complemento avverbiale (*Adverbiale, adverbiale Bestimmung*), detto anche semplicemente avverbiale, aggiunge informazioni relative al luogo (*Ortsangaben*), al tempo (*Zeitangaben*), alla modalità, alla causa e alla finalità di un determinato evento. Nella Tabella 2.3 una sintesi dei tipi di avverbiale, con indicazione dei pronomi interrogativi con i quali possono essere chiesti e i pronomi con i quali possono essere sostituiti

Tipo di avverbiale	Specificità semantiche	Pronome interrogativo	Pronome
avverbiale di luogo	di stato	wo?	dort, hier, da
	di movimento	wohin? woher?	dorthin, dorthier
avverbiale di tempo	puntuale	wann?	dann
	durativo	wie lange?	so lange
	iterativo	wie oft	so oft
avverbiale di modo	modale	wie? auf welche Weise?	so
	graduabile	wie sehr?	so
	comitativo (di compagnia)	mit wem?	mit + pronome personale
	strumentale	womit?	damit
avverbiale di causa	causale	warum? weshalb?	deshalb, deswegen
	condizionale	in welchem Fall?	in diesem Fall

	finale	wozu? zu welchem Zweck?	dazu
	concessivo	trotz was?	trotzdem

**Tabella 2.3:** Tipi di avverbiale con indicazione dei pronomi interrogativi e i pronomi (cfr. Pittner 2016: 109)

La funzione di avverbiale può essere svolta da vari tipi di sintagmi, come nell'esempio (74): un sintagma preposizionale (74a), un sintagma avverbiale (74b), un sintagma aggettivale (74c), un sintagma nominale all'accusativo (74d), un sintagma nominale al genitivo (74e), una frase (74f). In base alla valenza del singolo verbo, gli avverbiali possono fungere sia da complemento di verbo (*Ergänzung*), sia da circostanziale (*Angabe*):

(74)

- a) **Auf einer Treppe** saßen Logan und Storm. (deTenTen20)  
auf einer Treppe = PP (Ergänzung)
- b) **Hier oben** steht ja „Deine Meinung zählt“. (deTenTen20)  
Hier oben = AdvP (Ergänzung)
- c) Excel rechnet also **falsch** oder wir haben eine fehlerhafte Formel eingegeben. (deTenTen20) (Angabe)  
falsch = AdjP
- d) Genießen sie **den NRW Tag**. (deTenTen20)  
den NRW Tag = Akkusativ-NP (Ergänzung)
- e) Und **eines Tages** kamen zwei sehr bekannte Menschen zu ihm. (deTenTen20)  
eines Tages = Genitiv-NP (Angabe)
- f) Einige von ihnen haben Bauchweh, **während andere Fieber haben**. (deTenTen20)  
während andere Fieber haben = Satz (Angabe)

#### *Predicativo*

Il predicativo (*Prädikativum*) si trova con verbi copulativi (detti anche verbi copulari), con verbi estimativi e con verbi appellativi.

Il predicativo indica una proprietà riferita al soggetto e all'oggetto della frase. Con i verbi copulativi (*sein*, *bleiben*, *werden*) il predicativo si riferisce al soggetto e insieme al verbo costituisce il predicato. Può assumere varie forme, come mostrato nell'esempio (75):

(75)

- a) Ich bin **müde**. (deTenTen20)  
müde = AdjP
- b) Er wird **Politiker**. (deTenTen20)  
Politiker = Nominativ-NP
- c) Die Amerikaner sind **außer sich**. (deTenTen20)  
außer sich = PP

d) Wir bleiben **ruhig**. (deTenTen20)  
 ruhig =AdvP

L'elemento predicativo è parte integrante della funzione predicativa<sup>47</sup>.

Il predicativo che invece indica una proprietà indicata dall'oggetto della frase si trova con verbi estimativi che esprimono un giudizio, come *finden*, o appellativi, come *nennen*, *halten für*, *bezeichnen*, che indicano un processo di denominazione.

Qui di seguito un esempio con *halten für* (76):

(76)  
 Ich halte es für **eine sehr gute Idee**. (deTenTen20)

Esiste anche il predicativo libero, che specifica uno stato temporaneo del soggetto (77) e dell'oggetto (78):

(77) Wir kamen **müde** zum Abendessen nach Hause. (deTenTen20)

(78) Nach dem Schwimmen: Weißweine **kühl** stellen. (deTenTen20)

A livello sintattico il predicativo libero non può essere distinto dall'avverbiale di modo. La distinzione avviene solo a livello semantico: il predicativo libero si riferisce allo stato del soggetto o dell'oggetto, mentre l'avverbiale al verbo.

Il predicato in alcune grammatiche viene considerato un costituente sintattico (cfr. Di Meola 2014: 118). In questo volume non si seguirà questa impostazione, dato che il predicato non può essere spostato senza modificare il tipo di frase; inoltre, è molto difficile, se non in taluni casi impossibile, applicare il test di domanda e pronominalizzazione.

### *Attributo*

Gli attributi (*Attribute*) sono parte di un *Satzglied* e sono dunque chiamati anche *Satzgliedteile* o *Gliedteile*. Di norma non possono essere spostati da soli, ma sono legati all'elemento a cui si accompagnano, un sostantivo o un aggettivo. Vi sono vari tipi di attributo (79):

(79)  
 Adjektiv-Attribut: **schöne** Frau;  
 PP-Attribut: die Zeitung **auf dem Sofa** ist vom Opa;  
 Genitiv-Attribut: die Freundin **meiner Schwester**;  
 Attributsatz: die Frau, **die dort geparkt hat**;  
 Apposition: Onkel **Dirk**.

<sup>47</sup> Gaeta lo esemplifica mediante il test di sostituzione (2017: 205).

### 2.1.4 Ordine delle parole a livello di frase

La frase in tedesco è relativamente libera e può essere adattata alle esigenze comunicative. Le regole per descrivere l'ordine delle parole a livello di frase sono pertanto complesse e un buon punto di partenza per la sua analisi è dato dall'identificazione del sintagma verbale. Le parti che compongono il sintagma verbale sono spesso distanti tra di loro; si dice infatti che formino una parentesi frasale (*Satzklammer*), chiamata anche *Verbalklammer* (parentesi verbale).

La *Verbalklammer* è composta da un verbo finito, coniugato (*Vorverb*) che rappresenta l'apertura della parentesi (la parentesi sinistra, *linke Klammer*), ed eventualmente da un verbo infinito (*Nachverb*), che costituisce la chiusura della parentesi (la parentesi destra, *rechte Klammer*). La frase viene quindi divisa in tre campi: il campo preposto (*Vorfeld*), che precede il *Vorverb*; il campo interposto (*Mittelfeld*), compreso tra il *Vorverb* e il *Nachverb*; il campo posposto (*Nachfeld*), che segue il *Nachverb*. Nella Tabella 2.4 l'esemplificazione della suddivisione della frase *Sie hat ihm eine Zigarre geschenkt*:

Vorfeld	Vorverb (linke Klammer)	Mittelfeld	Nachverb (rechte Klammer)	Nachfeld
Sie	hat	ihm eine Zigarre	geschenkt	

**Tabella 2.4:** Suddivisione di una frase in campi

Questo modello di descrizione della frase, chiamato anche *topologisches Satzmodell*, risalente a Drach (1937) e sviluppato poi ulteriormente, si è dimostrato molto utile per descrivere le regole che determinano l'ordine delle parole a livello di frase.

Il verbo finito assume diverse posizioni e determina il tipo di frase: se occupa la prima posizione (*Verb-Erststellung*), si tratta di una frase interrogativa totale (senza pronomi/avverbio interrogativo) o imperativa (Tabella 2.5).

Vorfeld	Vorverb (linke Klammer)	Mittelfeld	Nachverb (rechte Klammer)	Nachfeld
	Suchen	Sie etwas?		
	Mach	deine Hausaufgaben!		

**Tabella 2.5:** Articolazione di una frase interrogativa e imperativa in campi

Se il verbo finito occupa la seconda posizione (*Verb-Zweitstellung*), la frase è dichiarativa o interrogativa parziale (introdotta da un pronome/avverbio interrogativo) (Tabella 2.6):

Vorfeld	Vorverb (linke Klammer)	Mittelfeld	Nachverb (rechte Klammer)	Nachfeld
Sie	hat	ihm eine Zigarre	geschenkt	
Was	hat	sie ihm	geschenkt?	

**Tabella 2.6:** Suddivisione di una frase dichiarativa e interrogativa in campi

Per distinguere una frase interrogativa, imperativa e dichiarativa è molto importante anche l'intonazione: nella frase interrogativa è ascendente (nello scritto è presente un punto interrogativo); nella frase imperativa l'intonazione è discendente (nello scritto viene indicata da un punto esclamativo); nella frase dichiarativa l'intonazione è discendente (nello scritto è segnalata da un punto fermo).

Se il verbo finito occupa l'ultima posizione della frase (*Verb-Endstellung*) (Tabella 2.7), si tratta una frase secondaria (nella tabella sotto si trova nel *Nachfeld*).

Vorfeld	Vorverb (linke Klammer)	Mittelfeld	Nachverb (rechte Klammer)	Nachfeld
Ich	bin	in die Fotocommunity	gekommen,	weil ich hier Inspiration finde.

**Tabella 2.7:** Suddivisione di una frase complessa in campi

Le frasi secondarie prendono avvio da una congiunzione subordinante (*subordinierende Konjunktion*) o da un altro elemento subordinante.

Alcune indicazioni generali:

- il *Vorverb* può essere preceduto da un solo costituente sintattico. Statisticamente è più frequente il soggetto, ma il *Vorfeld* può essere occupato anche da altri costituenti sintattici, come l'oggetto (diretto e indiretto) e dagli avverbiali;
- nel *Mittelfeld* possono stare vari costituenti (sia per tipo che per numero);
- il *Nachfeld* può ospitare al massimo due costituenti sintattici; di solito è occupato dalle frasi secondarie.

I principali tipi di frasi in tedesco sono cinque, caratterizzati non solo dalla posizione del verbo coniugato, ma anche dall'intento comunicativo (*Sprechhandlung*): la frase dichiarativa (*Aussagesatz*); la frase interrogativa (*Fragesatz*), che a sua volta si distingue in *Ergänzungsfragesatz* (*w-Fragesatz*) ed *Entscheidungsfragesatz* (*Ja-Nein-Frage*); la frase imperativa (*Imperativsatz*); la frase esclamativa (*Exklamativsatz*); la frase ottativa (*Wunschsatz*).

Nelle frasi dichiarative, nelle frasi interrogative del tipo (*w-Fragesatz*), nelle frasi esclamative (*Wie schön ist die Landschaft!*) e nelle frasi secondarie introdotte dai verbi del dire (*Verben des Sagens*) e del pensare (*Verben des Denkens*), che hanno la forma di frasi principali dichiarative, il verbo coniugato si trova in seconda posizione. Nelle frasi interrogative del tipo *Entscheidungsfragesatz*, nelle

frasi imperative ed esclamative (*hat der aber Pech gehabt!*), nonché in barzellette e racconti (*sagt Klein-Erna zu seiner Oma*), il verbo coniugato si trova invece in prima posizione.

In tutte le secondarie introdotte da una congiunzione o da un elemento subordinante, il verbo coniugato si trova in posizione finale. Vi sono anche frasi secondarie usate come frasi autonome (*dass der da alles weiß!*).

### 2.1.5 Frasi principali e secondarie

Le frasi possono essere semplici o complesse. Se sono semplici, sono autonome e possono stare da sole. Il periodo complesso, invece, è caratterizzato da più frasi, che si possono trovare sullo stesso livello gerarchico (coordinazione) o essere in rapporto gerarchico di subordinazione (Tabella 2.8).

Nel primo caso si usa il termine “paratassi” (*Parataxe* o *Satzreihung*). Le frasi si trovano in un rapporto di coordinazione tra di loro, o per mezzo di congiunzioni coordinanti e avverbi (polisindeto) o senza alcun collegamento (asindeto). Le principali congiunzioni coordinanti<sup>48</sup> sono *und*, *oder*, *aber*, *sondern*, *denn*. *Und* ha un significato additivo, *oder* ha un significato disgiuntivo, *aber* e *sondern* hanno un significato avversativo e *denn* un significato causale. In taluni casi, l’ordine delle frasi coordinate può essere liberamente modificato; in altri, quando vi è una sequenza temporale o temporale-causale, no. Frasi coordinate possono essere a loro volta subordinate a un’altra frase.

Nel secondo caso si usa il termine “ipotassi” (*Hypotaxe* o *Satzgefüge*): le frasi sono in un rapporto di subordinazione tra di loro. Le frasi che sono in forma di subordinazione (le frasi secondarie) sono definite *Nebensätze*. Per ogni frase subordinata la rispettiva frase sovraordinata è chiamata *Matrixsatz* (cfr. Busch/Stenschke 2018: 136). Le frasi sovraordinate possono essere a loro volta subordinate ad altre frasi (si parla di secondaria di primo grado, di secondo grado ecc.). Le frasi che, invece, non sono dipendenti da altre vengono definite frasi principali (*Hauptsätze*). L’ordine tra principale e subordinata può essere invertito. Nella Tabella 2.8 un riassunto schematico della frase complessa:

Frase complessa			
Paratassi		Ipotassi	
asindeto	polisindeto	frase principale/frase sovraordinata	frase secondaria o subordinata

**Tabella 2.8:** La frase complessa

Oltre al fattore dell’autonomia o della dipendenza da altre frasi, anche la posizione del verbo coniugato rappresenta un elemento importante in tedesco. Come già esemplificato nel paragrafo 2.1.4, in base a questo criterio le frasi

<sup>48</sup> La coordinazione non avviene solo a livello di frase, ma anche tra costituenti.

principali si distinguono da quelle secondarie; le prime, infatti, presentano il verbo coniugato in seconda posizione<sup>49</sup>, le seconde in ultima posizione<sup>50</sup>.

Le frasi secondarie si possono distinguere in base alla forma, alla funzione<sup>51</sup> e alla semantica (cfr. Busch/Stenschke 2018: 136-137).

Per quanto concerne la forma, le secondarie possono essere introdotte da un elemento o non presentare alcun elemento.

Se introdotte da un elemento, questo può presentare forme diverse in base alle quali si distingue il tipo di frase secondaria. L'elemento può essere una congiunzione subordinante, e in questo caso si ha un *Konjunktionalsatz* (80). Può trattarsi anche di un pronome o avverbio relativo; in questo caso si ha un *Pronominalnebensatz*, che può essere un *Relativsatz* (81), se il pronome è relativo (*Relativum*), o un *abhängiger Fragesatz*, se il pronome è interrogativo (*w-Interrogativum*) (82):

(80) Wir wollen so verhindern, **dass das Coronavirus in unser Land eingeschleppt wird**. (deTenTen20)

(81) Bächli Bergsport bietet ein sehr breites Bekleidungsassortiment, **das sowohl funktionellen als auch modischen Ansprüchen gerecht wird**. (deTenTen20)

(82) Keine Ahnung, **warum das so ist**. (deTenTen20)

Le frasi introdotte da un elemento hanno il verbo all'ultimo posto (*Verbendstellung*). Le frasi che non vengono introdotte da alcun elemento possono presentare il verbo al secondo posto (*uneingeleitete Verbzweitnebensätze*), il verbo al primo posto (*uneingeleitete Verberstnebensätze*) o avere un verbo non finito.

Nel caso di *uneingeleitete Verbzweitnebensätze* (il verbo è al secondo posto) si tratta di frasi secondarie che seguono i verbi del dire e del pensare (83) o i loro rispettivi sostantivi (84).

(83) Regisseur Mittendrein **meinte**, Regensburg verfüge über ein gutes Theater. (deTenTen20)

(84) ÖVP und SPÖ vertraten dagegen **die Meinung**, die Steuer sei nicht mehr notwendig. (deTenTen20)

Nel caso di *uneingeleitete Verberstnebensätze* (il verbo è al primo posto) sono solitamente frasi secondarie condizionali o concessive (85).

(85) Aber so sind wir nun mal, **regnet es**, wird gemeckert. (deTenTen20)

Le *uneingeleitete Verbzweitnebensätze* e le *uneingeleitete Verberstnebensätze* sono frasi secondarie finite, ossia hanno il verbo finito.

49 Eccezioni sono le interrogative totali, le frasi imperative e alcune esclamative.

50 Eccezioni sono le secondarie non introdotte da congiunzione: *Er glaubt er kriegt jetzt einen TablettPC* (deTenTen).

51 Per la suddivisione in base alla funzione sintattica cfr. Bianco/Mollica (2024) e la *Dudengrammatik* (cfr. Duden 2022: 160s.).

Le frasi che non hanno un verbo finito sono definite *satzwertige Infinitivphrasen* o *satzwertige Partizipphrasen*. Possono essere trasformate in una frase secondaria introdotta da un elemento subordinante (86).

(86)

Ich kann dir also nicht versprechen, **niemals einen Fuß in dein Land zu setzen**. (deTenTen20)  
 = Ich kann dir also nicht versprechen, **dass ich niemals einen Fuß in dein Land setzen werde**.

Le frasi secondarie possono svolgere una funzione sintattica all'interno della frase sovraordinata (1.) oppure non svolgere alcuna funzione sintattica (*weiterführender Nebensatz*, *Satzrelativsatz*) (2.).

(1) Frasi secondarie con funzione sintattica

Nel caso delle frasi secondarie con funzione sintattica, si deve ulteriormente distinguere se (a.) hanno funzione di costituente sintattico (*Gliedsatz*) o (b.) se fanno parte di un costituente sintattico (*Gliedteilsatz*).

a. Frasi secondarie con funzione di costituente sintattico

Le frasi secondarie possono svolgere vari ruoli sintattici nel periodo complesso. In particolare, possono avere funzione di soggetto, oggetto, predicativo e avverbiale (la funzione dipende dal verbo della frase principale).

Se una frase secondaria svolge la funzione di soggetto (87), si trova al posto del sintagma nominale al nominativo. Deve rispondere alla domanda *wer? was?* e può essere sostituita da un pronome.

(87) **Dass sie kommt**, war schon klar. (deTenTen20)

Se la frase è posposta, può essere preannunciata da un pronome; di norma si tratta del pronome *es*<sup>52</sup> (88):

(88) **Es** war schon klar, **dass sie kommt**. (deTenTen20)

Il pronome *es*, in questa funzione, viene definito correlato (*Korrelat*). Il termine “correlato” si riferisce a elementi, per lo più anaforici, dal valore lessicale nullo, che uniscono la frase principale con le loro secondarie<sup>53</sup> (cfr. Bianco/Mollica 2024). I correlati svolgono principalmente una funzione cataforica (ossia rimandano a qualcosa che viene dopo). L'occorrenza di *es*, in quanto

52 In alcuni casi può esserci anche *das* (cfr. Di Meola 2014: 122).

53 Per approfondimenti sui correlati e sulla loro occorrenza (facoltativa vs. obbligatoria) si rimanda a Bianco/Mollica (2024).

correlato, dipende sia dal verbo della frase principale, sia dalla posizione della frase subordinata (cfr. Bianco/Mollica 2024).

Se una secondaria svolge la funzione di oggetto, in caso di oggetto diretto (89) può occupare il posto del sintagma nominale all'accusativo e risponde alla domanda *wen? was?*

(89) Ich glaube, **dass dies Früchte tragen kann.** (deTenTen20)

Come si evince dall'esempio (89), la secondaria non è preannunciata da alcun correlato. Le secondarie che svolgono la funzione di oggetto (diretto) e sono posposte possono, quindi, non essere preannunciate da un correlato. La presenza di un correlato dipende, anche in questo caso, dal verbo della frase sovraordinata.

La secondaria con la funzione di oggetto indiretto può occupare il posto del sintagma nominale al dativo (90) (*wem? was?*) o del sintagma nominale al genitivo (91) (*wessen?*):

(90) So kann er die Fische geben, **wem er will.** (deTenTen20)

(91) Sie war sich dessen bewusst, **später einmal Königin zu sein.** (deTenTen20)

La secondaria con funzione di oggetto preposizionale può sostituire il sintagma preposizionale (92):

(92) Wir rechnen damit, **dass die Auswahl bis Ende August abgeschlossen werden kann.** (deTenTen20)

La presenza di un avverbio preposizionale (come *damit, darauf, davon* ecc.) con funzione di correlato dipende, anche in questo caso, dal verbo e dalla posizione della frase subordinata. In (93) un esempio con *damit*:

(93) Wir rechnen **damit, dass die Auswahl bis Ende August abgeschlossen werden kann.** (deTenTen20)

= **Dass die Auswahl bis Ende August abgeschlossen werden kann, damit** rechnen wir.

In (94) un esempio di secondaria con funzione predicativa:

(94) Die Liebe ist, **was** für immer bleibt, bis in alle Ewigkeit. (deTenTen20)

Le frasi secondarie che svolgono la funzione di avverbiale aggiungono informazioni spaziali, temporali, modali ecc. Le domande che si possono utilizzare sono varie (*wo? wann? warum? wozu?*). Le frasi secondarie locative completano con informazioni di tipo spaziale (*wo? wohin? woher?*); quelle temporali forniscono

informazioni di tipo temporale (*wann?*), laddove la relazione cronologica può essere di contemporaneità, posteriorità e anteriorità (cfr. Tabella 2.9):

Relazione	Congiunzioni
contemporaneità	während, indem, als, solange, sobald, sowie, sooft, wie wenn
posteriorità	nachdem, seit, seitdem, als, sobald, sowie, wenn
anteriorità	bis, bevor, ehe

**Tabella 2.9:** Relazioni delle congiunzioni di tipo temporale

Altre frasi secondarie sono (cfr. Di Meola 2014: 124-125; Duden 2022: 167s.): le frasi modali, che offrono informazioni sul modo di svolgimento di un’azione/un evento (*indem, wie, dadurch dass, sodass, als ob*); le frasi avversative, in cui è presente contrasto o incompatibilità (*während, wogegen, wohingegen, wo, wenn*); le frasi causali, che indicano la causa/il motivo (*weil, da, zumal, dafür dass*); le frasi finali, che esprimono lo scopo (ossia l’obiettivo) di una determinata azione (*dass, damit*); le frasi consecutive, che indicano la conseguenza di un determinato fatto o il suo effetto (*dass, sodass, als dass*); le frasi condizionali, che esprimono la condizione necessaria affinché si verifichi un determinato fatto (*wenn, falls, sofern, soweit*); le frasi concessive, che descrivono una causa in contrasto con un effetto in apparenza inatteso (*obwohl, obgleich, obschon, auch wenn* ecc.)<sup>54</sup>.

Il ruolo sintattico delle secondarie dipende dal verbo della frase sovraordinata, ossia dalla sua valenza. Nel caso le frasi secondarie siano richieste dalla valenza del verbo sono definite frasi complemento (in tedesco *Komplementsätze* o *Ergänzungssätze*)<sup>55</sup>, nel caso non lo siano sono definite frasi circostanziali (in tedesco *Angabesätze* o *Adverbialsätze*) (cfr. Pittner 2016: 121 e Bianco/Mollica 2024). Le frasi secondarie circostanziali, anche se sono facoltative da un punto di vista grammaticale, svolgono un ruolo molto importante dal punto di vista comunicativo.

#### b. Frasi secondarie parte di un costituente sintattico

Se le frasi secondarie sono parte di un costituente, sono collegate di norma a un sostantivo, come mostrano gli esempi (95-97) e svolgono funzione attributiva. Si possono trovare vicino al sostantivo o essere staccate da esso (nel *Nachfeld*). Non possono mai comparire nel *Vorfeld* senza il sostantivo di riferimento.

54 Come può essere evinto da questo breve excursus sui tipi di secondarie, vi sono sia congiunzioni sinonimiche (esprimono le stesse relazioni logico-semantiche), sia congiunzioni che possono essere usate per esprimere più relazioni logico-semantiche (per esempio, *während* può essere sia temporale sia avversativa). Inoltre, alcune congiunzioni coincidono morfologicamente con altre classi di parole (*seit, bis, während*, ad esempio, sono anche preposizioni).

55 Per un elenco delle frasi complemento in base alla frequenza d’uso cfr. Bianco/Mollica (2024).

- (95) Eine kühle Flasche Bier, **die er vom Wege in seinem selbst angelegten Keller lagerte**, soll schon diese Wanderung wert gewesen sein. (deTenTen20)
- (96) Neben ihm stand immer die Flasche Bier, **die er alle fünf Minuten zum Mund führte**. (deTenTen20)
- (97) Bei der Siegerehrung hat mein Vater eine 2 l Flasche Bier bekommen, **die als Pokal gewertet wird**. (deTenTen20)

Tra le frasi attributive vi sono sia le relative determinative, ossia quelle che contengono informazioni sull'antecedente precisandone il senso, sia le frasi relative appositive, che forniscono informazioni supplementari su un antecedente già identificato in maniera univoca (cfr. Di Meola 2014: 123).

(2) Frasi secondarie senza funzione sintattica

In questo caso le frasi secondarie queste non possono essere chieste o pronominalizzate. Di norma si trovano nel *Nachfeld* (98):

- (98) Mit einer Starthand von Ass König in Karo kam dann die Dame, **was mich natürlich sofort zum Chipleader gemacht hat**. (deTenTen20)

Le frasi secondarie occupano tutti i campi della frase: *Vorfeld*, *Mittelfeld* (più raramente) e *Nachfeld*. Nella Tabella 2.10 alcuni esempi di analisi topologica<sup>56</sup> delle frasi complesse (esempi 97-100). Si ricorda che l'analisi parte sempre dalla frase principale (*Hauptsatz*):

Vorfeld	Vorverb (linke Klammer)	Mittelfeld	Nachverb (rechte Klammer)	Nachfeld
Eine kühle Flasche Bier, die er vom Wege in seinem selbst angelegten Keller lagerte,	soll	schon diese Wanderung wert	gewesen sein.	
Neben ihm	stand	immer die Flasche Bier, die er alle fünf Minuten zum Mund führte.	#	
Bei der Siegerehrung	hat	mein Vater eine 2 l Flasche Bier	bekommen,	die als Pokal gewertet wird.
Mit einer Start-hand von Ass König in Karo	kam	dann die Dame,	#	was mich natürlich sofort zum Chipleader gemacht hat.

**Tabella 2.10:** Esempi di analisi topologica di frasi complesse

<sup>56</sup> Per approfondimenti si veda il volume Blühdorn/Foschi (2014).

Le frasi secondarie possono essere analizzate in maniera più specifica in una fase successiva (Tabella 2.11):

Vorfeld	Vorverb (linke Klammer)	Mittelfeld	Nachverb (rechte Klammer)	Nachfeld
Eine kühle Flasche Bier, die er vom Wege in seinem selbst angelegten Keller lagerte,	soll	schon diese Wanderung wert	gewesen sein.	
	die	er vom Wege in seinem selbst angelegten Keller	lagerte	

**Tabella 2.11:** Esempi di analisi topologica di frasi complesse, inclusa l'analisi della secondaria

## 2.2 Sintassi con i corpora

In questo secondo paragrafo sono presentati e discussi due casi studio. Il primo prende in esame elementi polifunzionali della lingua tedesca, come il lessema *ganz*, utilizzabile come attributo (in tal caso, viene declinato) o come avverbio e intensificatore (non viene declinato). *Ganz* verrà analizzato nel *web corpus German Web 2020* e nel corpus di tedesco scritto DeReKo, utilizzando le rispettive funzionalità. I risultati verranno messi a confronto per formulare prime ipotesi sulle combinazioni più frequenti di *ganz* come spunto di riflessione per ulteriori approfondimenti e per analisi in altri corpora. Il secondo caso studio si concentra sugli avverbi pronominali, in particolare quelli con *wo-*. Gli avverbi pronominali possono avere varie funzioni, per esempio possono essere usati come elementi forici, deittico-situazionali e correlati (cfr. Ravetto 2022). Analizzando il corpus di tedesco *German Web 2020*, il focus sarà posto su *worin*. Si osserva, in particolare, se sono presenti forme disintegrate/discontinue e quale funzione è predominante.

### 2.2.1 Caso studio I: l'elemento polifunzionale *ganz*

Come altri elementi linguistici della lingua tedesca, *ganz* è polifunzionale, ossia svolge varie funzioni sintattiche (Glück 2000: 537). La *Dudengrammatik* (cfr. Duden 2016: 1304) sottolinea la sua versatilità: l'aggettivo *ganz*<sup>57</sup> può essere, infatti, usato come attributo, avverbio e intensificatore. Quando usato come attributo, concorda in caso, numero e genere con il sostantivo che lo

<sup>57</sup> *Ganz* viene inserito tra gli aggettivi quantitativi (cfr. Duden 2022: 781); per le sue caratteristiche semantiche particolari non può essere usato in forma comparativa (Trost 2006: 192). Per i suoi significati si veda anche Trost (2006: 198). Per una spiegazione degli usi di *ganz* si veda Flinz (2021a: 284-291).

segue; come intensificatore (Breindl 2009) o avverbio non si modifica, ossia non viene flesso. Proprio per la sua polifunzionalità *ganz* costituisce spesso un problema per gli apprendenti e i traduttori. Gli apprendenti tendono a invertirne le caratteristiche: lo declinano quando è intensificatore e non lo declinano quando è usato come attributo (Flinz 2021a: 282). I traduttori, invece, hanno difficoltà a identificarne il significato corretto se usato come intensificatore; infatti, *ganz*, che può modificare sintagmi aggettivali, avverbiali, verbali, preposizionali, ma mai nominali, può avere un significato *intensivierend-steigernd* o *abschwächend-abstufend* in base a diversi fattori, tra cui la semantica dei costituenti a cui fa riferimento (la funzione *abschwächend* è combinata con una valutazione soggettiva), l'accento (nella funzione *steigernd* la particella intensificatrice *ganz* è accentata, mentre in quella *abschwächend* no, si vedano anche Pusch 1981; Rinas 2013), il contesto, che svolge un ruolo importante per determinare la funzione (cfr. Flinz/Moroni 2020: 125).

Né le grammatiche per apprendenti *DaF* né le risorse lessicografiche sono di aiuto per distinguere gli usi e le traduzioni di *ganz*, mentre i corpora possono dare un supporto concreto (Flinz 2022; Flinz/Mell 2022).

Anche a livello teorico molte delle indicazioni sono basate su esempi costruiti *ad hoc*, mentre solo recentemente sono stati utilizzati corpora per identificare caratteristiche specifiche di *ganz*, anche con un approccio della grammatica delle costruzioni (Flinz/Moroni in pubb.).

Nel presente caso studio utilizzeremo il corpus di testi CMC estratti da Internet *German Web 2020* (20.999.598.683 token) e il corpus di tedesco scritto DeReKo-2023-1 con l'obiettivo di analizzare con quali elementi si combina preferibilmente *ganz*, che funzione svolge e che tipo di sintagmi forma.

Per analizzare *German Web 2020* verrà utilizzata la funzionalità *Word-Sketch*, con la quale possono essere estratti in maniera automatica profili di parola<sup>58</sup> (*Word Sketches*); per analizzare DeReKo invece la *Kookkurenzanalyse* per l'estrazione di profili sintagmatici.

Analizzando il corpus *German Web 2020*, si evince immediatamente che le occorrenze di *ganz* come avverbio e intensificatore (14.455.764) sono superiori a quelle di *ganz* come attributo (7.455.797). Per quanto riguarda il suo uso come avverbio/intensificatore, gli elementi con cui si combina preferibilmente sono riassunti nella Figura 2.1:

58 Dalle liste dei profili di parola si possono evincere la frequenza della combinazione e il *logdice*, che indica quando forte, e quindi tipica, è la combinazione estratta.

<b>besonder</b>	372,689	10.4	...
ganz besondere			
<b>ander</b>	373,040	10.3	...
ganz andere			
<b>einfach</b>	354,904	10.2	...
ganz einfach			
<b>normal</b>	310,668	10.1	...
ganz normal			
<b>klar</b>	306,741	10.0	...
ganz klar			
<b>schön</b>	327,488	9.9	...
ganz schön			
<b>gut</b>	489,053	9.7	...
ganz gut			
<b>neu</b>	239,722	9.6	...
ganz neue			
<b>sicher</b>	177,752	9.3	...
ganz sicher			
<b>genau</b>	168,248	9.1	...
ganz genau			

**Figura 2.1:** *WordSketch* di *ganz* (avverbio/intensificatore) in *German Web 2020*

Dalle prime dieci combinazioni emerge che l'elemento con cui si combina preferibilmente *ganz* è un aggettivo. Nella Tabella 2.13 sono stati inseriti anche alcuni esempi (usi più frequenti):

DeTenTen2020	Co-occorrenza	Usi più frequenti	Tipo di sintagma
ganz	besonder	ein <b>ganz</b> besonderes Erlebnis	sintagma nominale
	ander	ein <b>ganz</b> anderes Kaliber	sintagma nominale
	einfach	ein <b>ganz</b> einfacher Trick <b>ganz</b> einfach beantworten	sintagma nominale, sintagma aggettivale
	normal	der <b>ganz</b> normale Wahnsinn, <b>ganz</b> normal verhalten	sintagma nominale, sintagma aggettivale
	klar	eine <b>ganz</b> klare Kaufempfehlung, <b>ganz</b> klar sagen	sintagma nominale, sintagma aggettivale
	schön	<b>ganz</b> schön nerven, <b>ganz</b> schön finden <b>ganz</b> schön ins Schwitzen	sintagma aggettivale, sintagma aggettivale, sintagma aggettivale
	gut	<b>ganz</b> gut klappen <b>ganz</b> gut in Griff haben	sintagma aggettivale, sintagma aggettivale
	neu	eine <b>ganz</b> neue Dimension <b>ganz</b> neu entdecken	sintagma nominale, sintagma aggettivale
	sicher	<b>ganz</b> sicher wiederkommen	sintagma aggettivale
	genau	<b>ganz</b> genau hinschauen <b>ganz</b> genau unter der Lupe nehmen	sintagma aggettivale, sintagma aggettivale

**Tabella 2.13:** Prime 10 co-occorrenze di *ganz*, uso e tipo di sintagma

Anche estraendo i dati da DeReKo-2023-1 (10.764.300.838 forme), le occorrenze di *ganz* come avverbio e intensificatore (6.958.287) sono maggiori di quelle come attributo (3.527.765). Le co-occorrenze più frequenti e tipiche sono (Tabella 2.14):

DeReKo	Co-occorrenza	Modello sintagmatico	Tipo di sintagma
ganz	anders	<b>ganz</b> anders als	sintagma avverbiale
	oben	<b>ganz</b> oben auf der Wunschliste	sintagma avverbiale
	klar	<b>ganz</b> klar	sintagma aggettivale
	einfach	<b>ganz</b> klar	sintagma aggettivale
	schön	<b>ganz</b> schön anstrengend	sintagma aggettivale
	zu schweigen	<b>ganz</b> zu schweigen	sintagma verbale
	nebenbei	<b>ganz</b> nebenbei	sintagma avverbiale
	besonders	<b>ganz</b> besonders	sintagma avverbiale
	sicher	<b>ganz</b> sicher	sintagma aggettivale
	normal	<b>ganz</b> normal	sintagma aggettivale

**Tabella 2.14:** Prime 10 co-occorrenze di *ganz*, modello sintagmatico e tipo di sintagma

Confrontando i dati estratti dal *web corpus* e dal corpus di tedesco scritto emerge che, nonostante la base dati sia diversa (CMC vs. scritto), vi sono molte combinazioni in comune (*ganz* + *ander/anders, klar, einfach, schön, besonderes/besonders, sicher, normal*). Nella lista delle prime 10 co-occorrenze in *German Web 2020* compaiono anche *ganz gut, ganz neu, ganz genau*, fatto che potrebbe essere collegato al tipo di corpus. *Ganz genau, ganz gut* sono usate spesso come espressioni valutative. In DeReKo occorrono anche *ganz oben, ganz zu schweigen, ganz nebenbei*.

Per quanto riguarda il tipo di sintagma, in entrambi i corpora domina il sintagma aggettivale. Il sintagma nominale è presente solo in *German Web 2020*.

Per quanto concerne invece l'uso di *ganz* come attributo, in *German Web 2020* sono state trovate 7.455.797 occorrenze; i sostantivi con cui si accompagna più frequentemente sono:

1. *Welt: auf der ganzen Welt* (sintagma preposizionale);
2. *Tag: den ganzen Tag* (sintagma nominale);
3. *Familie: für die ganze Familie* (sintagma preposizionale);
4. *Reihe: eine ganze Reihe* (sintagma nominale);
5. *Leben: ein ganzes Leben* (sintagma nominale);
6. *Jahr: das ganze Jahr über* (sintagma preposizionale);
7. *Körper: den ganzen Körper* (sintagma nominale);
8. *Land: im ganzen Land* (sintagma preposizionale);
9. *Menge: eine ganze Menge* (sintagma nominale);
10. *Weile: eine ganze Weile* (sintagma nominale).

In DeReKo-2023-1 sono state trovate 3.527.765 occorrenze di *ganz* come attributo. I sostantivi con cui si combina più frequentemente sono:

1. *Welt: auf der ganzen Welt* (sintagma preposizionale);

2. *Reihe: eine ganze Reihe* (sintagma nominale);
3. *Tag: den ganzen Tag* (sintagma nominale);
4. *Land: im ganzen Land* (sintagma preposizionale);
5. *Familie: für die ganze Familie* (sintagma preposizionale);
6. *Menge: eine ganze Menge* (sintagma nominale);
7. *Leben: ihr/sein ganzes Leben* (sintagma nominale);
8. *Jahr: das ganze Jahr über* (sintagma preposizionale);
9. *Weile: eine ganze Weile* (sintagma nominale);
10. *Region: in der ganzen Region* (sintagma preposizionale).

Anche nelle combinazioni di *ganz* come attributo emergono somiglianze tra i due corpora, forse ancora maggiori: nelle 10 combinazioni più frequenti e più tipiche vi sono praticamente gli stessi sostantivi (*ganz* + *Welt*, *Reihe*, *Tag*, *Land*, *Familie*, *Menge*, *Leben*, *Jahr*, *Weile*). I dati differiscono solo per *ganz* + *Körper* in *German Web 2020* e *ganz* + *Region* in *DeReKo*. Anche gli usi delle combinazioni, come si evince dagli esempi, sono simili. In entrambi i corpora prevalgono i sintagmi nominali e preposizionali.

Con questo principio di analisi si è potuto osservare che:

- in entrambi i corpora prevale l'uso di *ganz* come avverbio e come intensificatore;
- l'uso di *ganz* come attributo è molto simile in entrambi i corpora: si combina frequentemente con gli stessi sostantivi andando a formare prevalentemente gli stessi sintagmi nominali e preposizionali;
- anche l'uso di *ganz* come avverbio e come intensificatore è molto simile nei due corpora, nonostante emergano alcune differenze in parte riconducibili al tipo di dati (CMC e scritto).

Ulteriori analisi potrebbero essere condotte con altri tipi di corpora, per esempio con *FOLK* per analizzare il tedesco parlato, oppure con *InterCorp* per concentrarsi sulla traduzione.

### 2.2.2 Caso studio II: avverbi pronominali in *wo-* in *German Web 2020*. Il caso di *worin*

Gli avverbi pronominali<sup>59</sup>, chiamati anche avverbi preposizionali (in tedesco *Pronominaladverbien* o *Präpositionaladverbien*), sono formati da un avverbio e da una preposizione, e hanno la funzione di sostituire un sintagma preposizionale o addirittura un'intera frase (Duden 2022: 803). Sono una classe chiusa, essendo formati prevalentemente dagli avverbi *da-*, *wo-* o, in misura minore, da *hier-* e dalle preposizioni *an*, *auf*, *aus*, *bei*, *durch*, *für*, *gegen*, *hinter*, *in/ein*, *mit*, *nach*, *neben*, *über*,

59 Si vedano anche Pasch et al. (2003); Negele (2012); Ravetto (2022).

*um, unter*. Se la preposizione inizia per vocale, si inserisce anche *-r* (*da-r-auf*); nella lingua parlata il *dar-* viene spesso ridotto a *dr-* (Duden 2022: 803).

Ravetto (2022) analizza gli avverbi in *da(r)-*, mettendo in evidenza come vi sia un continuum tra forme più integrate e forme meno integrate (disintegrate, nel senso di discontinue cfr. Negele 2012), con una tendenza anche alla separazione tra l'avverbio *da* e la preposizione (come in *Der kann zwei Köpfe größer sein, da habe ich keine Angst vor* o in *wir hoffen auf Eure Einsendungen – da freuen wir uns drauf* in cui l'elemento deittico *da* è in posizione iniziale, mentre la preposizione è alla fine della frase) e con forme raddoppiate (come in *Wat soll ich denn dadarauf*)<sup>60</sup>. Queste forme discontinue si trovano soprattutto nella lingua colloquiale e in alcuni dialetti del nord della Germania (Spiteller 1972: 12). Proprio le forme meno integrate costituiscono spesso un problema per gli apprendenti; ciononostante, non vengono discusse nelle grammatiche per apprendenti *DaF* (cfr. Ravetto 2022: 50-53).

Nel presente caso studio il focus è sugli avverbi pronominali in *wo(r)-*, che, secondo la letteratura, presentano minore possibilità di combinazione rispetto a *da(r)* (cfr. Ravetto 2022: 23).

Come gli altri avverbi pronominali, gli avverbi pronominali in *wo(r)-* possono svolgere diverse funzioni:

- la funzione forica, quando si riferiscono al contesto precedente (anaforico) o al contesto successivo (cataforico). In questo caso, sono fondamentali per la coesione (Andorno 2003; Ravetto 2021, 2022);
- la funzione deittico-situazionale, quando si riferiscono a entità del contesto situazionale;
- la funzione di correlato (Mollica 2010; Schwabe 2013), che anticipa frasi successive. Dipendono dalla valenza di verbi, nomi, aggettivi e si riferiscono a frasi secondarie.

Estrapolando le occorrenze delle varie forme in *German Web 2020* (Tabella 2.15) si evince che *wobei* è l'avverbio più frequente, mentre *woneben* e *wozwischen* sono quelli meno frequenti:

Pronominaladverb	Frequenza assoluta	Frequenza ppm
wobei	2.683.425	127,78
wodurch	612.224	29,15
worauf	475.329	22,64
womit	458.794	21,85
wozu	351.340	16,73
wofür	323.921	15,43

60 Gli esempi citati sono presi dal corpus *German Web 2020*. Cfr. anche Flinz/Ravetto in prep.

wonach	322.275	15,35
woran	295.354	14,06
wovon	257.075	12,24
worin	155.870	7,42
worum	142.258	6,77
worüber	100.934	4,81
woraus	82.875	3,95
wogegen	40.986	1,95
worunter	21.852	1,04
wovor	21.168	1,01
worein	1056	0,05
wohinter	779	0,04
woneben	158	0,01
wozwischen	80	<0,01
wowider	13	<0,01

**Tabella 2.15:** Frequenza assoluta e normalizzata (ppm) degli avverbi pronominali in *German Web 2020*

In particolare, è evidente la frequenza nettamente più alta di *wobei* rispetto agli altri avverbi. Il numero di occorrenze può però essere spiegato con il fatto che *wobei*, come anche *damit*, può ricorrere nel corpus anche con la funzione di congiunzione subordinante (99).

(99) Im Durchschnitt enthält 100 Gramm Fisch ungefähr 20 Gramm Eiweiß, **wobei** sich der genaue Wert natürlich von Art zu Art unterscheidet. (deTenTen20)

*Woneben*, invece, ha un numero molto basso di occorrenze e in DeReKo, in linea con i dati analizzati da Ravetto (2022), è usato quasi esclusivamente come pronome relativo (100).

(100) Dann deutete sie auf ein Waschbecken, **woneben** auch Desinfektionsmittel war. (deTenTen20)

Solo in due casi viene utilizzato come pronome interrogativo (101):

(101) In Word 2000 erscheinen standardmäßig in der Taskleiste separate Schaltflächen je Dokument. Das lässt sich nicht ändern (im Gegensatz zu Excel 2000). Zitat: **Woneben?** Wenn (abhängig von genannter Option) für jede Datei ein eigenes Fenster geöffnet ist, steht dir in Word ein Schließfeld in der Menüleiste zur Verfügung, um nacheinander die einzelnen Fenster und schließlich das Programm zu beenden. (deTenTen20)

Anche *wozwischen* è attestato come pronome relativo (102) e pronome interrogativo (103):

(102) Nun war klar, **wozwischen** die DDR-Bürger in einigen Tagen die Wahl haben würden: die D-Mark 1:1 oder Kollaps. (deTenTen20)

(103) **Wozwischen** musst du dich denn entscheiden? (deTenTen20)

Rispetto ad altri studi basati su dati di tedesco scritto (si ricorda che *German Web 2020* è un corpus di testi estratti da Internet, quindi tipici della comunicazione in rete) le occorrenze di *wohinter*, *woneben*, *wozwischen*, *wowider* non sono così rare o assenti.

*Worin* e *worein* sono presenti entrambe; si distinguono per il fatto che, mentre *worin* indica stato in luogo (104), *worein* segnala un movimento (105) (cfr. Duden 2016: 592):

(104) Man braucht in einen Dampfkessel von 17 1/2 Fuß Länge und 3 1/2 Fuß Durchmesser, **worin** täglich 15 bis 18 Hektoliter Wasser verdampft werden. (deTenTen20)

(105) liebe ratte, komm zu mir, gerne spiele ich mit dir, bind dir engleinsflügel um, trag dich ins panoptikum, **worein** oft die kinder gehn, und wann die dich fliegen sehn, rufen alle, alle aus: so ne große fledermaus. (deTenTen20)

Per quanto riguarda la funzione sintattica, gli avverbi pronominali in *wo(r)*- si distinguono da quelli *da(r)*- e *hier*- poiché utilizzabili esclusivamente come avverbi interrogativi o pronomi relativi<sup>61</sup> (Ravetto 2022: 24).

In questo caso studio il focus verte su *worin*; in particolare, si analizzano le forme standard (*worin*) e le forme che si trovano nella posizione intermedia del continuum (Tabella 2.16).

Le domande di ricerca sono le seguenti: sono presenti forme discontinue in *German Web 2020* e come si comportano a livello sintattico?

Proadv.	Frequenza assoluta	Esempio
worin	155.870	Ergänzend hat Gabi kürzlich noch einen weiteren Blog-Eintrag verfasst, in dem sie ein Buch vorstellt, <b>worin</b> es um das Verhältnis von Theorie und Empirie geht
wo rin	28	Der minimalistisch verspielte Einschlag unterliegt oft einen Hypnotischen drive <b>wo rin</b> sich die housigen Sounds gehör verschaffen.
woworin	0	

61 Si veda anche l'uso come *weiterführender Relativsatz*, che allude al contenuto di un intero enunciato (cfr. Holler 2005).

wo worin	4	„Sind wir schon in Paris?“ Ich forderte ihn auf, sich „Jedermann“ anzusehen, <b>wo worin</b> gezeigt wird, wie schwer den reichen Leuten das Sterben fällt.
wodrin	350	Edeldame Rotfeld bat sich an eine Mappe zu führen, <b>wodrin</b> alle Berichte und Informationen gebündelt werden.
wo drin	2.095	Hallo Anke, ich denke, hier im Forum gibt es so um die 50 Beiträge, <b>wo drin</b> steht, wie man ausschleicht und zwei Auswertungen, wo es auch noch mal explizit beschrieben ist
wowodrin	0	
wo wodrin	0	

**Tabella 2.16:** Forme disintegrate dell'avverbio pronominale *worin* in *German Web 2020*

Alla prima domanda di ricerca può essere risposto che le forme disintegrate non sono molte e sono poco frequenti. Sono state identificate, infatti, solo le seguenti forme: *wo rin*, *wo worin*, *wo drin*. A livello sintattico hanno prevalentemente il ruolo di pronomi relativo, in misura minore quello di pronomi interrogativo.

Tra le forme discontinue e disintegrate la forma *wo drin* è quella più frequente. L'analisi di un suo campione casuale di 50 occorrenze ha messo in evidenza che, anche in questo caso, il ruolo sintattico dominante assunto è quello di pronomi relativo (106):

(106) Unten rechts erschien ein Balken **wo drin** stand "Level 80 erreicht".  
(deTenTen20)

Sono presenti pochi casi (3) in cui l'avverbio viene usato come pronomi interrogativo (107):

(107) Wann soll man Babys baden und **wo drin**? (deTenTen20)

Per quanto riguarda la funzione, invece, prevale quella forica, riferita soprattutto al contesto precedente.

Riassumendo, il caso studio qui esaminato ha mostrato come l'uso e la funzione di una forma, *worin*, possa variare in base al tipo di corpus esaminato. Nello specifico, è emerso che le combinazioni con *wo(r)*- sono diverse in un *web corpus* rispetto a un corpus scritto: sono presenti forme pressoché assenti in altri dati così come forme discontinue/disintegrate. Il ruolo assunto è sempre quello di pronomi relativo o pronomi interrogativo. La funzione dominante è quella forica.

## 2.3 Attività ed esercizi<sup>62</sup>

### 2.3.1 Classi di parola

Teoria → Capitolo 2.1.1

**1.** Cercare in un corpus a piacere un nome/sostantivo (per esempio *Fisch*): identificare il singolare e il plurale. Analizzando 10 frasi, fare ipotesi sul genere e sul caso (se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX).

**2.** Cercare in un corpus a piacere un aggettivo (per esempio *schön*): identificare la presenza o meno dell'articolo e riflettere sulla marca flessiva. Analizzando 10 frasi, identificare eventuali forme comparative e superlative (se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX).

**3.** Cercare in un corpus a piacere un aggettivo (per esempio *schön*): analizzare la sua posizione all'interno della frase (se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX).

Rispondere alle seguenti domande:

- L'aggettivo è in funzione attributiva o predicativa?
- Se è in funzione attributiva, è declinato?
- Se è in funzione predicativa è declinato?

**4.** Cercare in un corpus a piacere le occorrenze di *gestrig* (se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX). Che cosa si può dedurre? È usato in funzione attributiva? È usato in funzione predicativa?

**5.** Cercare in un corpus a piacere le occorrenze di *quitt* (se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX). Che cosa si può dedurre? È usato in funzione attributiva? È usato in funzione predicativa?

**6.** Cercare in un corpus a piacere la forma *schön* (se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX). Oltre alla funzione attri-

---

<sup>62</sup> Per la scelta della forma linguistica “a piacere” usare come supporto la parte teorica.

butiva e predicativa, trovare esempi in funzione di avverbio e di intensificatore. Discutere gli esempi.

**7.** Cercare in un corpus a piacere un articolo (per esempio *der*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, ricerca CQL, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX. Qui di seguito un supporto per la ricerca CQL:

- ricerca CQL [word="der"] [tag="N.\*"]: analizzare e commentare 10 occorrenze (*sentence*)
- ricerca CQL [word="der"] [] [tag="N.\*"]: analizzare e commentare 10 occorrenze (*sentence*)
- ricerca CQL [word="der"] [] [] [tag="N.\*"]: analizzare e commentare 10 occorrenze (*sentence*)
- ricerca CQL [word="der"] [] [] [] [tag="N.\*"]: analizzare e commentare 10 occorrenze (*sentence*)

Aumentare il numero degli slot vuoti [] a piacere e riflettere. Quanti slot posso inserire? Fare la medesima procedura con altri articoli determinativi (*die, das*) o altri tipi di articolo, come indeterminativo (*ein*), negativo (*kein*), possessivo (*mein*), dimostrativo (*dieser, jener, solcher*), interrogativo (*welcher*), indefinito universale (*jener, mancher, irgendein*).

**8.** Cercare in un corpus a piacere un articolo. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, ricerca CQL ([tag="ART.\*"]), scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Classificare 10 occorrenze in articolo determinativo (*der*), indeterminativo (*ein*), negativo (*kein*), possessivo (*mein*), dimostrativo (*dieser, jener, solcher*), interrogativo (*welcher*), indefinito universale (*jener, mancher, irgendein*).
- Ipotizzare poi anche il caso e il genere.

**9.** Cercare in un corpus a piacere la forma *der*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Ipotizzare se la forma è usata come articolo o come pronome.
- Di che tipo di pronome si tratta?

**10.** Cercare in un corpus a piacere il verbo *sein*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, ricerca CQL ([lemma="sein"] [tag="V.\*"]), scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX. Analizzare 20 occorrenze.

- Indicare per quale tempo verbale è utilizzato (*Perfekt* e *Plusquamperfekt*).
- Vi sono occorrenze in cui il verbo essere è usato per fare il passivo di stato (*Zustandspassiv = sein + Partizip II*)?

- Fare la ricerca anche con *haben* (ricerca CQL [lemma=“haben”][tag=“V.\*”]) e analizzare il tempo verbale (*Perfekt* e *Plusquamperfekt*).
- Fare la ricerca anche con *werden* (ricerca CQL [lemma=“werde”][tag=“V.\*”]) e indicare se è utilizzato per fare il futuro (*Futur I* e *Futur II*), *Konjunktiv* (con *würde*), *Vorgangspassiv* (*werden* + *Partizip II*).

**11.** Cercare in un corpus a piacere un verbo modale (*können*, *dürfen*, *müssen*, *sollen*, *wollen*, *mögen* + *Infinitiv*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l’opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Estrarre 20 occorrenze e identificare il tempo verbale.
- Riflettere sul tipo di modalità: deontica (obbligo, necessità, permesso ecc.) o epistemica (probabilità).

**12.** Cercare in un corpus a piacere la categoria grammaticale avverbi. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, ricerca CQL [tag=“ADV.\*”], scegliere l’opzione di visualizzazione per *sentence* e filtrare tramite GDEX. Analizzare 20 occorrenze.

- Identificare gli avverbi di tempo (a che domanda rispondono? *wann?* *wie lange?* *wie oft?*) e di luogo (a che domanda rispondono? *wo?* *wohin?* *woher?*)
- Procedere con altri tipi di avverbio (modali, causali). Analizzare 10 occorrenze, andando a focalizzare il sintagma avverbiale e la sua posizione nella frase (in quale campo si trova?).

**13.** Cercare in un corpus a piacere una delle seguenti preposizioni: *mit*, *von*, *an*, *auf* e *zu*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l’opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Analizzare 20 occorrenze e identificare il caso che reggono.
- Analizzare 10 occorrenze con *an* e indicare se la preposizione è utilizzata come preposizione temporale o locale.

**14.** Analizzare 20 occorrenze di *trotz* (o *wegen*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l’opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Identificare il caso retto. Genitivo? Dativo?

**15.** Analizzare 20 occorrenze di una delle seguenti preposizioni: *an*, *auf*, *hinter*, *in*, *neben*, *über*, *unter*, *vor*, *zwischen*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l’opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Identificare il caso che reggono e riflettere se si tratta di stato in luogo o di movimento.
- 16.** Analizzare 20 occorrenze di *und* (o *oder*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.
- Riflettere se la congiunzione coordina frasi del solito livello o porzioni di frase.
- 17.** Cercare una delle seguenti congiunzioni subordinanti: *dass* e *ob* (frasi dichiarative e interrogative), *wenn, sobald, weil, als, wenn* (avverbiali); *nachdem, während, als, wenn* (temporali); *wenn, falls, im Fall(e)* (condizionali); *weil, da, wo* (causali); *obwohl, wengleich* (concessive) e *als, als dass* (comparative). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.
- Estrarre 20 occorrenze e individuare il verbo della frase secondaria. In quale posizione si trova?
- 18.** Cercare una delle seguenti particelle modali: *aber, auch, bloß, denn, doch*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.
- Estrarre 20 occorrenze e dire in che tipo di frase si trovano (affermative, interrogative, imperative)?
  - Analizzare la posizione in base al verbo coniugato: in quale campo si trovano?
- 19.** Le unità linguistiche usate come particelle, possono essere spesso usate anche in altre classi di parola, come aggettivi (*rubig, eigentlich*), avverbi (*schon, vielleicht*), congiunzioni (*denn, aber*) ecc. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.
- Estrarre 20 occorrenze di *aber* e individuare se l'unità linguistica è congiunzione o particella.
- 20.** Estrarre 20 occorrenze della particella focalizzante *nur*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.
- Individuare l'elemento focalizzato.
- 21.** Estrarre 20 occorrenze della particella intensificatrice *sehr*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Dopo aver individuato il verbo coniugato, indicare in quale campo si trovano e l'elemento a cui fanno riferimento.

**22.** Cercare le seguenti interiezioni: *mmb*, *na ja*, *br*, *gell*, *aua*, *igitt*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Dopo aver individuato il verbo coniugato, indicare in quale campo si trovano.

### 2.3.2 Sintagmi

Teoria → Capitolo 2.1.2 Sintagmi

**1.** Estrarre le occorrenze di un sostantivo (per esempio *Schildkröte*, *Hund*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX. Analizzarne 10. Riflettere sul tipo di sintagma, individuando la testa:

Es. *die Schildkröte* → SN; *eine grüne Schildkröte* (SN); *die Schildkröte meines Bruders* (SN) ...;

- Indicare se nei sintagmi vi sono attributi e di che tipo:
  - attributo aggettivale (*eine grüne Schildkröte*);
  - attributo al genitivo (*die Schildkröte meines Bruders*);
  - attributo preposizionale (*die Schildkröte von meinem Bruder*);
  - frase attributiva (*die Schildkröte, die wir gestern im Garten gefunden haben*);
- Riflettere sull'ordine Testa – Complemento: quale ordine è presente?

**2.** Estrarre le occorrenze di una preposizione a piacere (per esempio *auf*, *unter*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX. Analizzarne 10. Individuare il sintagma preposizionale.

Es. *auf dem Tisch*; *auf dem alten Tisch*; *auf dem Tisch* ...;

- Indicare la testa. Quali elementi compongono il sintagma preposizionale?
- Riflettere sull'ordine Testa – Complemento: quale ordine è presente?

**3.** Estrarre le occorrenze di un verbo pieno (*Vollverb*) a piacere (per esempio *trinken*, *baden*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX. Analizzarne 10.

- Ipotizzare quali posizioni nella frase sono obbligatorie e quali facoltative.
- Indicare la testa (il verbo coniugato, quindi anche l'ausiliare!)

- Il verbo di trova in una frase principale o secondaria?
  - Riflettere sull'ordine Testa – Complemento: quale ordine è presente?
4. Estrarre le occorrenze di un aggettivo a piacere (per esempio *groß, sauer*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX. Analizzarne 10.
- Riflettere sul sintagma aggettivale. Quali altri elementi vi sono?
  - L'aggettivo è usato in funzione attributiva o predicativa?
  - Riflettere sull'ordine Testa – Complemento: quale ordine è presente?
5. Estrarre le occorrenze di un avverbio a piacere (per esempio *unten, dort*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX. Analizzarne 10.
- Riflettere sul sintagma avverbiale. Quali altri elementi vi sono?
  - Riflettere sull'ordine Testa – Complemento: quale ordine è presente?
6. Estrarre le occorrenze di un elemento linguistico a piacere e analizzare una frase. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.
- Individuare il verbo;
  - Individuare gli altri costituenti che compongono la frase e classificarli in sintagmi;
  - Svolgere il test di domanda per ogni sintagma;
  - Svolgere il test di pronominalizzazione per ogni sintagma;
  - Svolgere il test di permutazione;
  - Svolgere il test di topicalizzazione;
  - Dopo i vari test indicare quali sono i costituenti sintattici.

### 2.3.3 Funzioni sintattiche nella frase semplice

Teoria → Capitolo 2.1.3

1. Estrarre le occorrenze di un elemento linguistico a piacere e analizzare una frase. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.
- Individuare il verbo;
  - Individuare il soggetto (è al nominativo, concorda per numero e persona al verbo coniugato, risponde alla domanda *wer? was?*);

- Individuare se presente l'oggetto diretto (è all'accusativo, risponde alla domanda *wen? was?*);
- Individuare se presente l'oggetto indiretto (è al dativo, risponde alla domanda *wem? was?* O al genitivo risponde alla domanda *wessen?*);
- Individuare se presente l'oggetto preposizionale, che consiste in una preposizione fissata idiomaticamente dal verbo e da un sintagma nominale;
- Individuare se presente l'avverbiale (luogo, tempo, modo, causa). Come supporto per le domande usare la Tabella 2.4.

Elementi linguistici suggeriti:

- il verbo *untersuchen*;
- il verbo *grauen*. È presente il soggetto?
- il verbo *schenken*. È presente un oggetto indiretto?
- il verbo (*sich*) *erinnern*. È presente l'oggetto indiretto al genitivo o l'oggetto preposizionale?
- il verbo *geben*. Quali tipi di oggetto sono presenti (diretto, indiretto, entrambi)?
- il verbo *danken*. Quali tipi di oggetto sono presenti (diretto, indiretto, entrambi)?
- il verbo (*sich*) *schämen*. È presente l'oggetto indiretto al genitivo?
- il verbo *warten*. Individuare l'oggetto preposizionale;
- il verbo (*sich*) *freuen*. Individuare l'oggetto preposizionale;
- il verbo *wohnen*. Sono presenti avverbiali?
- il verbo *sein*. È presente un predicativo?
- il verbo *bleiben*. È presente un predicativo?
- il verbo *werden*. È presente un predicativo?
- il verbo *halten für*. È presente un predicativo?

### 2.3.4 Ordine delle parole a livello di frase

Teoria → Capitolo 2.1.4

**1.** Estrarre le occorrenze di un elemento linguistico a piacere e analizzare una frase. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Individuare il sintagma verbale, indicando la *Verbalklammer*. L'apertura della parentesi è il verbo coniugato; la chiusura eventualmente un verbo infinito. Riflettere sul tipo di frase (dichiarativa, interrogativa, imperativa);
- Se non è presente la chiusura della parentesi identificare la sua eventuale posizione;
- Annotare i tre campi (preposto, interposto, posposto);
- Individuare i costituenti sintattici;
- Fare ipotesi sul significato della frase.

### 2.3.5 Frasi principali e secondarie

Teoria → Capitolo 2.1.5

1. Estrarre le occorrenze di una delle seguenti congiunzioni coordinanti: *und, oder, aber, sondern, denn*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Identificare il verbo principale della frase introdotta dalla congiunzione. In quale posizione si trova?
- La coordinazione è a livello di frasi o di costituenti?

2. Estrarre le occorrenze di *dass*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Identificare il verbo principale della frase introdotta dalla congiunzione subordinante. In quale posizione si trova?
- Identificare la frase sovraordinata (*Matrixsatz*).
- Indicare in quale campo (preposto, interposto, posposto) si trova la frase secondaria.

3. Estrarre le occorrenze di un pronome relativo (*der, die, das ...*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, CQL, [word=","] [word="der"]), scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Identificare il verbo principale della frase introdotta dalla congiunzione subordinante. In quale posizione si trova?
- Identificare la frase sovraordinata (*Matrixsatz*).
- Indicare in quale campo (preposto, interposto, posposto) si trova la frase secondaria.

4. Estrarre le occorrenze di un pronome relativo il pronome è interrogativo (*warum, was, wo ...*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, CQL [word=","] [word="warum"]), scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Identificare il verbo principale della frase introdotta dalla congiunzione subordinante. In quale posizione si trova?
- Identificare la frase sovraordinata (*Matrixsatz*).
- Indicare in quale campo (preposto, interposto, posposto) si trova la frase secondaria.

5. Estrarre le occorrenze di un verbo del dire e del pensare (*meinen*) o di un loro sostantivo (*Meinung*). Se si usa il tool *Sketch Engine*, scegliere il corpus *German*

*Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Individuare l'eventuale presenza di una secondaria.
- Identificarne il verbo principale.
- Indicare in quale campo (preposto, interposto, posposto) si trova la frase secondaria.

**6.** Estrarre le occorrenze di una delle seguenti congiunzioni subordinanti: *während, indem, als, solange, sobald, sowie, sooft, wie wenn*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Identificare il verbo principale della frase introdotta dalla congiunzione subordinante. In quale posizione si trova?
- Identificare la frase sovraordinata (*Matrixsatz*).
- Indicare in quale campo (preposto, interposto, posposto) si trova la frase secondaria.

**7.** Estrarre le occorrenze di una delle seguenti congiunzioni subordinanti: *nachdem, seit, seitdem, als, sobald, sowie, wenn*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX.

- Identificare il verbo principale della frase introdotta dalla congiunzione subordinante. In quale posizione si trova?
- Identificare la frase sovraordinata (*Matrixsatz*).
- Indicare in quale campo (preposto, interposto, posposto) si trova la frase secondaria.

**8.** Estrarre le occorrenze di *während*. Se si usa il tool *Sketch Engine*, scegliere il corpus *German Web 2020*, usare *Concordance*, scegliere l'opzione di visualizzazione per *sentence* e filtrare tramite GDEX. Analizzare 10 frasi e discutere i risultati.

- Indicare se *während* ha il ruolo di congiunzione o preposizione. Motivare;
- Indicare, nel caso abbia il ruolo di congiunzione, che tipo di frase secondaria introduce: una frase secondaria temporale o avversativa. Indicare in quale campo (preposto, interposto, posposto) si trova la frase secondaria.

## Nota conclusiva

La descrizione di genesi e funzionamento di fenomeni osservabili ai vari livelli del sistema linguistico, oggetto di studio privilegiato della disciplina *DaF* d'ambito universitario, si fonda oggi inevitabilmente su procedure e prodotti della linguistica dei corpora. Il dialogo con la linguistica dei corpora all'interno del settore disciplinare *Lingua e Traduzione Tedesca*, già profilatosi diversi anni fa (*Deutsch als Fremdsprache – DaF*) (cfr. Foschi Albert/Hepp 2010), appare come una necessità non più rimandabile. Si è tentato in questo volume di compiere un passo in questa direzione, offrendo uno sguardo di sintesi sui risultati recenti della linguistica corpora e presentando la descrizione di alcuni fenomeni esemplari a livello di morfologia e sintassi del tedesco in una prospettiva sinergica tra la componente teorico-metodologica e quella pratica.



# Bibliografia

- Barbaresi, Adrien/Würzner, Kay-Michael (2014). For a fistful of blogs: Discovery and comparative benchmarking of republishable German content. In: *Workshop proceedings of the 12th edition of the KONVENS conference*. Hildesheim: Hildesheim University Press, 2-10. <https://d-nb.info/1126253383/34>
- Barbera, Manuel (2013). *Linguistica dei corpora e linguistica dei corpora italiana. Un'introduzione*. Milano: Qu.A.S.A.R.
- Baroni, Marco/Bernardini, Silvia (2004). BootCaT: Bootstrapping Corpora and Terms from the Web. In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*. Lisbon: European Language Resources Association (ELRA), 1313-1316.
- Baroni, Marco/Bernardini, Silvia (eds.) (2006). *WaCky! Working papers on the web as corpus*. Bologna: Gedit.
- Beißwenger, Michael (2013). *Das Dortmunder Chat-Korpus: ein annotiertes Korpus zur Sprachverwendung und sprachlichen Variation in der deutschsprachigen Chat-Kommunikation*. Online-Publikation auf dem Linguistik Server Essen (LINSE).
- Beißwenger, Michael/Imo, Wolfgang/Fladrich, Marcel/Ziegler, Evelyn (2019). <https://www.mocoda2.de>: a database and web-based editing environment for collecting and refining a corpus of mobile messaging interactions. In: *European Journal of Applied Linguistics* 7 (2), 333-344.
- Beißwenger, Michael/Lemnitzer, Lothar/Müller-Spitzer, Carolin (2022). *Forschen in der Linguistik. Eine Methodeneinführung für das Germanistik-Studium*. Paderborn: UTB.
- Beißwenger, Michael/Lüngen, Harald (2022). Korpora internetbasierter Kommunikation. In: M. Beißwenger/L. Lemnitzer/C. Müller-Spitzer: *Forschen in der Linguistik. Eine Methodeneinführung für das Germanistik-Studium*. Paderborn: UTB, 431-450.
- Beißwenger, Michael/Storrer, Angelika (2008). Corpora of Computer-Mediated Communication. In: A. Lüdeling/M. Kytö (eds.): *Corpus Linguistics. An International Handbook*. Volume 1. Berlin/New York: de Gruyter, 292-308.
- Betten, Anne/Du-nour, Miryam (1995). *Wir sind die Letzten. Fragt uns aus. Gespräche mit den Emigranten der dreißiger Jahre in Israel*. 1.-3. Aufl. Gerlingen: Bleicher (1995-1998). Neuauflage Gießen: Psychosozialverlag (2004).
- Betten, Anne/Flinz, Carolina/Leonardi, Simona (2023). Emigrantendeutsch in Israel: Die Korpora IS, ISW und ISZ der AGD/DGD. In: M. Kupietz/T. Schmidt: *Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache (CLIP)*. Band 11. Tübingen: Narr, 171-188.

- Bianco, Maria Teresa/Mollica, Fabio (2024). *La sintassi del verbo tedesco. La prospettiva contrastiva*. Roma: Carocci. In pubblicazione.
- Biber, Douglas et al. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Björkelund, Anders et al. (2014). The Extended DIRNDL Corpus as a Resource for Coreference and Bridging Resolution. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Island: European Language Resources Association (ELRA), 3222-3228. [http://www.lrec-conf.org/proceedings/lrec2014/pdf/891\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/891_Paper.pdf)
- Bloomfield, Leonard (1933/2001). *Language*. Beijing: Foreign Language Teaching and Research Press.
- Blühdorn, Hardarik (2017). Diskursmarker: Pragmatische Funktion und syntaktischer Status. In: H. Blühdorn/A. Deppermann/H. Helmer/T. Spranz-Fogasy (Hrsg.): *Diskursmarker im Deutschen. Reflexionen und Analysen*. Göttingen: Verlag für Gesprächsforschung, 311-336.
- Blühdorn, Hardarik/Foschi Albert, Marina (2014). *Leggere e comprendere il tedesco. Manuale per studenti e docenti in formazione*. Pisa: Pisa University Press.
- Bondi, Marina/Scott, Mike (eds.) (2010). *Keyness in texts*. (Studies in corpus linguistics 41). Amsterdam/Philadelphia: John Benjamins.
- Booij, Geert (2010). Compound construction: Schemas or analogy? A construction morphology perspective. In: S. Scalise/I. Vogel (Hrsg.): *Cross-Disciplinary Issues in Compounding* (Current Issues in Linguistic Theory. 311). Amsterdam/Philadelphia: Benjamins, 93-108.
- Bosco Coletsos, Sandra/Costa, Marcella (2013) (a cura di). *Italiano e tedesco. Questioni di linguistica contrastiva*. Alessandria: Dell'Orso.
- Brambilla, Marina/Flinz, Carolina (2020). Migrationsdiskurse in deutschen und italienischen Zeitungen: eine interlinguale datengeleitete Untersuchung. In: M. Brambilla/C. Flinz/R. Luppi (2020): *Deutsch im Vergleich: Texte und Diskurse* (Themenheft). *Annali. Sezione Germanica* 30 (2020). Napoli: UniorPress, 189-212. <http://www.serena.unina.it/index.php/aiongerm/article/view/8224/8945>
- Breindl, Eva (2009). Intensitätspartikel. In: L. Hoffmann (Hg.): *Handbuch der deutschen Wortarten*. Berlin/New York: de Gruyter: 397-422.
- Brinckmann, Caren/Kleiner, Stefan/Knöbl, Ralf/Berend, Nina (2008). German Today: an areally extensive corpus of spoken Standard German. In: *Proceedings 6th International Conference on Language Resources and Evaluation (LREC 2008)*. Marrakesch, Marokko: European Language Resources Association (ELRA), 3185-3191.
- Bubenhofner, Noah (2017). Kollokationen, n-Gramme, Mehrworteinheiten. In: K. Roth/M. Wengeler/A. Ziem (Hrsg.): *Handbuch Sprache in Politik und Gesellschaft*. Berlin/New York: de Gruyter, 69-93.

- Bühler, Hans et al. (1972). *Linguistik I. Lehr- und Übungsbuch zur Einführung in die Sprachwissenschaft*. 3. Auflage. Tübingen: Niemeyer.
- Burger, Harald (2015). *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Berlin: Erich Schmidt Verlag.
- Busch, Albert/Stenschke, Oliver (2018). *Germanistische Linguistik. Eine Einführung*. 4. Auflage. Tübingen: Narr.
- Busse, Dietrich/Teubert, Wolfgang (Hrsg.) (1994). Ist Diskurs ein sprachwissenschaftliches Objekt Zur Methodenfrage der historischen Semantik. In: D. Busse/F. Hermanns/W. Teubert: *Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik*. Opladen: Westdeutscher Verlag, 10-28.
- Bußmann, Hadumod (Hrsg.) (2007). *Lessico di Linguistica*. Traduzione italiana, adattamento e revisione sulla base della 3. Edizione originale, rivista ed ampliata a cura di Paola Cotticelli Kurras. Alessandria: Edizioni dell'Orso.
- Bußmann, Hadumod (Hrsg.) (2008). *Lexikon der Sprachwissenschaft*. 4. Aktualisierte und bibliographisch ergänzte Auflage. Stuttgart: Kröner.
- Cao, Yu Jie (2011). *X-GateX- (A Comparative Study of "X-Gate" and "X-" in Terms of Memetics)*. Masterarbeit, Huazhong: Huazhong University of Science and Technology.
- Čermáková, Ann et al. (2021). The International Comparable Corpus: Challenges in building multilingual spoken and written comparable corpora. In: *Research in Corpus Linguistics*, 9(1), 89-103. <https://doi.org/10.32714/ricl.09.01.06>
- Chomsky, Noam (1957). *Syntactic Structures*. Den Haag: Mouton.
- Cinato, Lucia (2020). *Voci di tedeschi in fuga. L'intervista autobiografica come contributo alla memoria collettiva*. Alessandria: Dell'Orso.
- Clyne, Michael (1981). *Deutsch als Muttersprache in Australien. Zur Ökologie einer Einwanderersprache*. In Zusammenarbeit mit dem Centre for Migrant Studies. Monash University. Deutsche Sprache in Europa und Übersee. Band 8. Wiesbaden: Franz Steiner Verlag.
- Cognola, Federica/Moroni, Manuela Caterina (2022). *Le particelle modali del tedesco. Caratteristiche formali, proprietà pragmatiche ed equivalenti funzionali in italiano*. Roma: Carocci.
- Costa, Marcella (2017). *Contrastività e traduzione. La morfologia valutativa in italiano e in tedesco*. Alessandria: Dell'Orso (Cultura tedesca 15).
- Deppermann, Arnulf/Hartung, Martin (2011). Was gehört in ein nationales Gesprächskorpus? Kriterien, Probleme und Prioritäten der Stratifikation des "Forschungs- und Lehrkorpus Gesprochenes Deutsch" (FOLK) am Institut für Deutsche Sprache (Mannheim). In: E. Felder/M. Müller/F. Vogel (Hrsg.): *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen*. Berlin/Boston: de Gruyter, 414-450.

- Deppermann, Arnulf/Schmidt, Thomas (2014). Gesprächsdatenbanken als methodisches Instrument der Interaktionalen Linguistik – Eine exemplarische Untersuchung auf Basis des Korpus FOLK in der Datenbank für Gesprochenes Deutsch (DGD2). In: C. Domke/C. Gansel (Hrsg.): *Korpora in der Linguistik – Perspektiven und Positionen zu Daten und Datenerhebung*, 4-17.
- Didakowski, Jörg/Geyken, Alexander (2014). From DWDS corpora to a German word profile – methodological problems and solutions. In: *OPAL – Online publizierte Arbeiten zur Linguistik* 2/2014, 39-47.
- Di Meola, Claudio (2014). *La linguistica tedesca. Un'introduzione con esercizi e bibliografia ragionata*. Nuova edizione rivista e ampliata. Roma: Bulzoni.
- Donalies, Elke (2002). *Die Wortbildung des Deutschen. Ein Überblick*. Tübingen: Narr.
- Donalies, Elke (2009). Stiefliches Geofaszintainment. Über Konfixtheorien. In: P.O. Müller (Hg.): *Studien zur Fremdwortbildung*. Hildesheim u.a.: Olms, 41-64.
- Drach, Erich (1937). *Grundgedanken der deutschen Satzlehre*. Frankfurt a. M.: Diesterweg.
- Dürscheid, Christa (2012). *Einführung in die Schriftlinguistik*. 4. Überarbeitet u. ergänzte Auflage. Göttingen: Vandenhoeck & Ruprecht.
- Eisenberg, Peter (2013a). *Grundriss der deutschen Grammatik. Das Wort*. Stuttgart/Weimar: Metzler.
- Eisenberg, Peter (2013b). *Grundriss der deutschen Grammatik. Der Satz*. Stuttgart/Weimar: Metzler.
- Engel, Ulrich (2009). *Deutsche Grammatik*. Neubearbeitung, 2. Aufl. München: Iudicium.
- Evert, Stefan (2009). Corpora and collocations. In: A. Lüdeling/M. Kytö (Hrsg.): *Corpus Linguistics*. Berlin/New York (Handbücher zur Sprach- und Kommunikationswissenschaft, 29), 1212-1248.
- Fandrych, Christian/Meißner, Cordula/Slavcheva, Adriana (2012). The GeWiss corpus: Comparing spoken academic German, English and Polish. In: T. Schmidt/K. Wörner (eds.): *Multilingual Corpora and Multilingual Corpus Analysis*. Hamburg Studies in Multilingualism 14. Amsterdam: John Benjamins, 319-338.
- Fandrych, Christian/Meißner, Cordula/Wallner, Franziska (2018). Das Potenzial mündlicher Korpora für die Sprachdidaktik. Das Beispiel GeWiss. In: *Deutsch als Fremdsprache* 55: 1, 3-13.
- Fandrych, Christian/Schmidt, Thomas/Wallner, Franziska/Wörner, Kai (Hrsg.) (2023). Zugänge zu mündlichen Korpora für DAF und DAZ: das ZuMult-Projekt. *Korpora Deutsch als Fremdsprache* 3(1) (2023). <https://kordaf.tu-journals.ulb.tu-darmstadt.de/issue/92/info/>
- Fandrych, Christian/Tschirner, Erwin (2007). Korpuslinguistik und Deutsch als Fremdsprache. In: *Deutsch als Fremdsprache* 44, 195-204.
- Fankhauser, Peter/Kupietz, Marc (2019). Analyzing domain specific word embeddings for a large corpus of contemporary German. In: International

- Corpus Linguistics Conference, Cardiff, Wales, UK, July 22-26, 2019. Mannheim: Leibniz-Institut für Deutsche Sprache (IDS). [https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/9117/file/Fankhauser\\_Kupietz\\_Analyzing\\_domain\\_specific\\_word\\_embeddings\\_2019.pdf](https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/9117/file/Fankhauser_Kupietz_Analyzing_domain_specific_word_embeddings_2019.pdf)
- Feilke, Helmuth (1996). *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt a. M.: Suhrkamp.
- Fillmore, Charles J. (1968). The Case for Case. In: E. Bach et al. (eds.). *Universals in Linguistic Theory*. New York: Rinehart & Winston, 1-88.
- Firth, John Rupert (1957). *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Flach, Susanne/Kopf, Kristin/Stefanowitsch, Anatol (2018). Skandale und Skandalchen kontrastiv: Das Konfix -gate im Deutschen und Englischen. In: R. Heuser/M. Schmuck (eds): *Stiefkinder der Onomastik*. Berlin: de Gruyter.
- Fleischer, Wolfgang/Barz, Irmhild (2012). *Wortbildung der deutschen Gegenwartssprache*. Berlin/New York: de Gruyter.
- Flinz, Carolina (2018). Der lexikographische Prozess bei Tourlex (ein deutsch-italienisches Fachwörterbuch zur Tourismussprache) für italienische DaF-Lerner. In: A. Klosa/A. Storrer/J. Taborek (Hrsg.) (2018): *Internetlexikographie und Sprachvermittlung*. Jahrbuch Lexicographica. Berlin: de Gruyter, 9-36.
- Flinz, Carolina (2019). Persuasionstrategien in deutschen rechtsorientierten Zeitungen. Eine korpuslinguistische Studie. In: F. Ricci Garotti/M.C. Moroni (Hrsg.): *Sprache und Persuasion. Sonderheft der Zeitschrift Linguistik Online* 97, 4/2019, 89-108. <https://bop.unibe.ch/linguistik-online/article/view/5597>
- Flinz, Carolina (2020). Vergleichbare Spezialkorpora für den Tourismus: eine Chance für den Fachsprachenunterricht. In: M. Hepp/K. Salzmann (Hrsg.) (2020): *Sprachvergleich in der mehrsprachig orientierten DaF-Didaktik. Theorie und Praxis*. Roma: Istituto Italiano di Studi Germanici, 133-151.
- Flinz, Carolina (2021a). Attributive Funktion und weitere Funktionen von ganz. Vorschläge für den DaF-Unterricht polyfunktionaler Wörter anhand von Korpora. In: C. Fandrych/M. Foschi Albert/M. Hepp/M. Thurmair: *Attribution in Text, Grammatik, Sprachdidaktik*. Berlin: Verlag Erich Schmidt, 275-301.
- Flinz, Carolina (2021b). Wahldiskurs in der Presse: eine korpusbasierte vergleichende Untersuchung zu den Parlamentswahlen in Deutschland und Italien. In: *APTUM. Zeitschrift für Sprachkritik und Sprachkultur*. 17. Jahrgang, 2021, Heft 01, 39-62.
- Flinz, Carolina (2021c). Korpora als primäre Quellen von Tourlex. In: J. Taborek/M. Piosok/M. Woznicka (Hrsg.) (2021): *Korpora in der Lexikographie*. Lexicographica. Series Maior. Berlin/Boston: de Gruyter, 57-83.
- Flinz, Carolina (2021d). Korpora in DaF und DaZ: Theorie und Praxis. In: C. Flinz/B. Hufeisen: *Korpora in DaF und DaZ: Theorie und Praxis. (Themenheft)*. *Zeitschrift für Interkulturellen Fremdsprachenunterricht. Didaktik und Methodik im Bereich*

- Deutsch als Fremdsprache*. Jahrgang 26, N.1 (April 2021), 1-43. <https://tjournals.ulb.tu-darmstadt.de/index.php/zif/>
- Flinz, Carolina (2022). #Verschwörungstheorien, #teorie del complotto. Eine deutsch-italienische Vergleichsstudie zum Corona-Impfverschwörungsdiskurs. In: *Muttersprache* 4/2022, 331-347.
- Flinz, Carolina/Gredel, Eva (2022). Eine kontrastive Korpusstudie zu Konfixderivaten: -phobie/-fobia und -manie/-mania im deutsch-italienischen Vergleich. In: L. Auteri et al.: *Wege der Germanistik in transkultureller Perspektive*. Band 6. Bern: Peter Lang Verlag, 365-380.
- Flinz, Carolina/Katelhön, Peggy (2019). Corpora nella didattica del tedesco come lingua straniera. Proposte per l'insegnamento del linguaggio specialistico del turismo. In: *ELLE*, 8(2),2019. Volume monografico, 323-348. <https://edizionicafoscarini.unive.it/it/edizioni4/riviste/elle/2019/2/corpora-nella-didattica-del-tedesco-come-lingua-st/>
- Flinz, Carolina/Luppi, Rita/Mell, Ruth M. (2024). Korpusgestützte Diskursanalyse und ihre Anwendbarkeit für den DaF-Unterricht am Beispiel des Klima-Diskurses. In: M. Pieklarz-Thien et. al.: *Forschungsgegenstände und ihre Verwendungsfelder*. In stampa.
- Flinz, Carolina/Mell, Ruth M. (2022). Über den Nutzen der Korpuslinguistik für den DaF-Unterricht beim Lehren und Lernen nicht-flektierbarer Wörter – eine Mikro-Studie zur Intensitätspartikel *ganz*. In: M. Pieklarz-Thien/S. Chudak (Hrsg.) (2022): *Wissenschaften und ihr Dialog in der Forschung zum Lehren und Lernen fremder Sprachen*. Göttingen: Vandenhoeck & Ruprecht, 211-230.
- Flinz, Carolina/Mell, Ruth M. (2023). Digitale Medien als Identitätsräume – Eine Analyse deutscher und italienischer Tweets zum Coronaleugner-Diskurs 2021. In: *Analisi Linguistica e Letteraria*. XXXI (2023), 5-24, <https://www.analisilinguisticaeletteraria.eu/index.php/ojs/article/view/542/458>
- Flinz, Carolina/Moroni, Manuela Caterina (in pubb.). Verwendungsspektrum der Intensitätspartikel *ganz* in narrativen biographischen Interviews. In: F. Mollica/A. Ziem (Hrsg.): *Intensivierung: Neue Impulse zu einem vielfältigen Phänomen*. Mannheim: Amades.
- Flinz, Carolina/Perkuhn, Rainer (2018). Wortschatz und Kollokationen in ‚Allgemeine Reisebedingungen‘. Eine intralinguale und interlinguale Studie. In: S. Krek et al. (eds.) (2018): *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Context*. Ljubljana University Press, Faculty of Arts: Ljubljana, 959-967. <https://euralex.org/publications/wortschatz-und-kollokationen-in-allgemeine-reisebedingungen-eine-intralinguale-und-interlinguale-studie-zum-fachsprachlich-lexikographischen-projekt-tourlex/>
- Foschi Albert, Marina/Hepp, Marianne (2010): Deutsch in Italien. In: H.J. Krumm/C. Fandrych/B. Hufeisen/C.Riemer (Hrsg.): *Deutsch als Fremd- und Zweisprache. Ein internationales Handbuch*. 2. Halbband. Berlin/New York: de Gruyter, 1693-1697.

- Foschi Albert, Marina/Hepp, Marianne (2012). Zusammenrückungen/ Conglomerati – ein vergleichender Blick auf die deutsche und italienische Substantivkomposition. In: L. Cinato et al.: *Intrecci di lingua e cultura*. Studi in onore di Sandra Bosco Colettos. Roma: Aracne, 119-146.
- Foth, Kilian A. et al. (2014). Because size does matter: The Hamburg Dependency Treebank. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Island: European Language Resources Association (ELRA), 2326-2333.
- Francis, W. Nelson/Kučera, Henry (1964). *A Standard Corpus of Present-Day Edited American English, for Use with Digital Computers*. Department of Linguistics, Brown University: Providence.
- Freddi, Maria (2019). *Linguistica dei corpora*. Roma: Carocci.
- Gaeta, Livio (2017). *Lineamenti di grammatica tedesca*. Roma: Carocci.
- Geyken, Alexander/Barbaresi, Adrien/Didakowski, Jörg/Jurish, Bryan/Wiegand, Frank/Lemnitzer, Lothar (2017). Die Korpusplattform des “Digitalen Wörterbuchs der deutschen Sprache” (DWDS). In: *Zeitschrift für Germanistische Linguistik*. Bd. 45, H2, 327-344.
- Glinz, Hans (1973). *Die innere Form des Deutschen. Eine neue deutsche Grammatik*. Sechste, durchgesehene Auflage mit zwei Faltafeln und einer Beilage. Bern/München: Francke.
- Glück, Helmut (2000). Komposition. In: H. Glück (Hrsg.): *Metzler-Lexikon Sprache* (4. Aufl.). Stuttgart: J.B. Metzler.
- Glück, Helmut/Rödel, Michael (Hrsg.) (2016). *Metzler Lexikon Sprache*. 5., aktualisierte und bearbeitete Auflage. Stuttgart: J.B. Metzler.
- Gredel, Eva/Flinz, Carolina (2020). Morphosyntax im deutsch-italienischen Vergleich: Eine kontrastive Fallstudie zu Wortbildungsprodukten in der internetbasierten Kommunikation der Online-Enzyklopädie Wikipedia. In: *Deutsche Sprache* 03/2020, 193-209.
- Hagemann, Jörg/Henle, Julia (2021). *Transkribieren nach GAT 2 (Minimal- und Basistranskript) – Schritt für Schritt*. Aktualisierte Version [https://drive.google.com/file/d/1fYms5RMu3-p4YFLg\\_ggkRLuOb6YEbYNO/view?usp=sharing](https://drive.google.com/file/d/1fYms5RMu3-p4YFLg_ggkRLuOb6YEbYNO/view?usp=sharing)
- Hausmann, Franz Josef (1984). Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen. In: *Praxis des neusprachlichen Unterrichts* 3, 4/1984, 395-406.
- Hundt, Marianne et al. (2007). *Corpus Linguistics and the Web*. Amsterdam/New York: Rodopi.
- Hüning, Matthias (2000). Monica en andere gates. Het ontstaan van een morfologisch procédé. In: *Nederlandse Taalkunde* 5/2, 121-132.
- Hunston, Susan (2008). Collection strategies and design decisions. In: A. Lüdeling/M. Kytö (eds.): *Corpus Linguistics. An International Handbook*. Volume 1. Berlin/New York: de Gruyter, 154-168.

- Imo, Wolfgang (2017). Interaktionale Linguistik und die qualitative Erforschung computervermittelter Kommunikation. In: M. Beißwenger (Hrsg.): *Empirische Erforschung internetbasierter Kommunikation*. Berlin: de Gruyter, 81-108.
- Imo, Wolfgang/Weidner, Beate (2018). Mündliche Korpora im DaF- und DaZ Unterricht. In: M. Kupietz/T. Schmidt (Hrsg.): *Korpuslinguistik*. Berlin/Boston: de Gruyter, 231-252.
- Joseph, Brian (1998). Diachronic morphology. In: A. Spencer/A. Zwicky (Hrsg.): *The handbook of morphology*. Oxford: Blackwell, 351-373.
- Kaiser, Julia (2018). Zur Stratifikation des FOLK-Korpus: Konzeption und Strategien. In: *Gesprächsforschung Online* 19, 515-552. <http://www.gespraechsforschung-online.de/fileadmin/dateien/heft2018/px-kaiser.pdf>
- Kaiser, Julia/Schedl, Eva (2021). Das Forschungs- und Lehrkorpus Gesprochenes Deutsch als Ressource für den handlungsorientierten DaF-Unterricht – Potentiale und Herausforderungen. In: C. Flinz/B. Hufeisen (Hrsg.): *Korpora in DaF und DaZ: Theorie und Praxis*. Themenausgabe der Zeitschrift für Interkulturellen Fremdsprachenunterricht, Jahrgang 26, Nummer 1 (April 2021), 45-83. <https://ojs.tu-journals.ulb.tu-darmstadt.de/index.php/zif/article/view/1113>
- Kemmer, Suzanne (2003). Schemas and lexical blends. In: H. Cuyckens et al. (Hrsg.): *Motivation in language*. (Current Issues in Linguistic Theory 243). Amsterdam/Philadelphia: Benjamins, 69-697.
- Kilgarriff, Adam (2009). Simple maths for keywords. In: M. Mahlberg/V. González-Díaz/C. Smith (eds.): *Proceedings of Corpus Linguistics Conference CL2009*. University of Liverpool, UK, July 2009. Liverpool: University of Liverpool. <https://www.sketchengine.eu/wp-content/uploads/2015/04/2009-Simple-maths-for-keywords.pdf>
- Kilgarriff, Adam/Grefenstette, Gregory (2003). Introduction to the special issue on the web as corpus. In: *Computational Linguistics* 29 (3), 333-347.
- Koch, Peter/Oesterreicher, Wulf (1994). Schriftlichkeit und Sprache. In: H. Günther/O. Ludwig (Hrsg.): *Schrift und Schriftlichkeit*. Berlin/New York: de Gruyter, 587-604.
- Konopka, Marek (2020). Grundlegende Aspekte der Wortbildung. In: M. Konopka/A. Wöllstein/E. Felder (Hrsg.): *Bausteine einer Korpusgrammatik des Deutschen*. Band 1. Heidelberg: Heidelberg, University Publishing, 29-63.
- Krekeler, Christian (2021). Korpustraining und datengestütztes Lernen im Wirtschaftsdeutsch-Unterricht. In: *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 26: 1, 161-195. <https://zif.tu-journals.ulb.tu-darmstadt.de/article/id/3326/>
- Krell, Gert (2020). *Klimadiskurs kontrovers: Zwischen Alarmstimmung und Verbarmlosung*. Frankfurt a. M.: Wochenschau Verlag.

- Krumm, Hans J./Fandrych, Christian/Hufeisen, Britta/Riemer, Claudia (Hrsg.) (2010). *Deutsch als Fremd- und Zweisprache. Ein internationales Handbuch*. 2. Halbband. Berlin/New York: de Gruyter.
- Kupietz, Marc/Lüngen, Harald/Kamocki, Pawel/Witt, Andreas (2018). The German Reference Corpus DeReKo: New Developments – New Opportunities. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA), 4353-4360.
- Lemnitzer, Lothar (2022). Korpora geschriebener Sprache. In: M. Beißwenger/L. Lemnitzer/C. Müller-Spitzer: *Forschen in der Linguistik. Eine Methodeneinführung für das Germanistik-Studium*. Paderborn: UTB, 411-420.
- Lemnitzer, Lothar/Diewald, Nils (2022). Abfrage und Analyse von Korpusbelegen. In: M. Beißwenger/L. Lemnitzer/C. Müller-Spitzer: *Forschen in der Linguistik. Eine Methodeneinführung für das Germanistik-Studium*. Paderborn: UTB, 374-390.
- Lemnitzer, Lothar/Zinsmeister, Heike (2006). *Korpuslinguistik: Eine Einführung*. Tübingen: Narr.
- Lemnitzer, Lothar/Zinsmeister, Heike (2015). *Korpuslinguistik: Eine Einführung*. 3. Auflage. Tübingen: Narr.
- Leonardi, Simona (2012). Scritture brevi in tedesco: forme e usi. In: F. Chiusaroli/F.M. Zanotto: *Scritture brevi nelle lingue moderne. Linguistica Zero*, 2, 2012, 95-112.
- Lüngen, Harald (2017). Das Deutsche Referenzkorpus DeReKo. Schriftkorpora der deutschen Gegenwartssprache am Institut für Deutsche Sprache in Mannheim. In: *Zeitschrift für germanistische Linguistik*. Bd. 45, Heft 1, 161-170.
- Lüngen, Harald/Kupietz, Marc (2017). CMC Corpora in DeReKo. In: P. Bański et al. (eds.): *Proceedings of the Workshop on Challenges in the Management of Large Corpora and Big Data and Natural Language Processing (CMLC-5+BigNLP 2017) including the papers from the Web-as-Corpus (WAC-XI) guest section*. Birmingham, 24 July 2017. Mannheim: Institut für Deutsche Sprache, 20-24.
- Lyons, John (1968). *Introduction to Theoretical Linguistics*. London: Cambridge University Press.
- Malloggi, Patrizio (2021). Zur Förderung der Kollokationskompetenz in der DaF-/Daz-Didaktik für den Tourismus anhand von DWDS-Korpora Korpora. In: *Deutsch als Fremdsprache* 1(1), 6-24. <https://doi.org/10.48694/tujournals-37>
- Manning, Christopher D./Schütze, Hinrich (2002). *Foundations of Statistical Natural Language Processing*. Cambridge (MA): MIT Press.
- McEnery, Tony/Wilson, Andrew (1996). *Corpus Linguistics. Edinburgh Textbooks in Empirical Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, Tony/Wilson, Andrew (2001). *Corpus Linguistics. Edinburgh Textbooks in Empirical Linguistics*. Edinburgh: Edinburgh University Press, 2. Edition.

- Morek, Miriam (2022). Gesprächsanalytische Transkription. In: M. Beißwenger/L. Lemnitzer/C. Müller-Spitzer (2022): *Forschen in der Linguistik. Eine Methodeneinführung für das Germanistik-Studium*. Paderborn: UTB, 391-410.
- Mukherjee, Joybrato (2002). *Korpuslinguistik und Englischunterricht*. Frankfurt am Main: Peter Lang.
- Ochs, Elinor (1979). Transcription as Theory. In: B. Schieffelin/E. Ochs (ed.): *Developmental Pragmatics*. New York, NY: Academic Press, 43-72.
- Perkuhn, Rainer/Keibel, Holger/Kupietz, Marc (2012). *Korpuslinguistik*. Paderborn: Fink.
- Pittner, Karin (2016). *Einführung in die germanistische Linguistik*. Darmstadt: WBG.
- Poesio, Massimo (2004). Coreference. In: *MATE Dialogue Annotation Guidelines-Deliverable 2.1*, 126-182. <http://www.andreasmengel.de/pubs/mdag.pdf>
- Pusch, Luise (1981). Ganz. In: Weydt, H. (Hg.): *Partikeln und Deutschunterricht*. Heidelberg: Groos, 31-43.
- Reineke, Silke/Deppermann, Arnulf/Schmidt, Thomas (2023). Das Forschungs- und Lehrkorpus für Gesprochenes Deutsch (FOLK). Zum Nutzen eines großen annotierten Korpus gesprochener Sprache für interaktionslinguistische Fragestellungen. In: A. Deppermann/C. Fandrych/M. Kupietz/T. Schmidt (Hrsg.): *Korpora in der germanistischen Sprachwissenschaft*. Berlin/Boston: de Gruyter, 71-102.
- Rinas, Karsten (2013). Nicht so GANZ einfach: Zur Semantik der ‚schillernden‘ Intensitätspartikeln ganz und docela. In: M. Nekula/K. Sichová/J. Valdrová (Hrsg.). *Bilingualer Sprachvergleich und Typologie: Deutsch-Tschechisch*. Tübingen: Groos, 223-244.
- Rossini Favretti, Rema (2000). Progettazione e costruzione di un corpus di italiano scritto – CORIS/CODIS. In: R. Rossini Favretti (a cura di): *Linguistica e informatica: corpora, multimedialità, percorsi di apprendimento*. Roma: Bulzoni, 39-56.
- Rossini Favretti, Rema (2001). La linguistica dei corpora in Europa: prospettive di analisi. In: *Lingua e Stile*. 2/2001, 367-382.
- Scherer, Carmen (2006). *Korpuslinguistik*. Heidelberg: Winter.
- Schmidt, Thomas (2022a). Daten und Metadaten. In: M. Beißwenger/L. Lemnitzer/C. Müller-Spitzer: *Forschen in der Linguistik. Eine Methodeneinführung für das Germanistik-Studium*. Paderborn: UTB, 249-258.
- Schmidt, Thomas (2022b). Korpora gesprochener Sprache. In: M. Beißwenger/L. Lemnitzer/C. Müller-Spitzer: *Forschen in der Linguistik. Eine Methodeneinführung für das Germanistik-Studium*. Paderborn: UTB, 421-430.
- Schmidt, Thomas/Schütte, Wilfried/Winterscheid, Jenny (2015). *cGAT. Konventionen für das computergestützte Transkribieren in Anlehnung an das Gesprächsanalytische Transkriptionssystem 2 (GAT2)*. [https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/4616/file/Schmidt\\_Schuette\\_Winterscheid\\_cGAT\\_2015.pdf](https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/4616/file/Schmidt_Schuette_Winterscheid_cGAT_2015.pdf)

- Scott, Mike (1997). PC analysis of key words – And key key words. In: *System* 25(2), 233-245.
- Scott, Mike/Tribble, Chris (2006). *Textual patterns: key words and corpus analysis in language education*. (Studies in corpus linguistics 22). Amsterdam/Philadelphia: John Benjamins.
- Sinclair, John (ed.) (1987). *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*. London-Glasgow: Collins ELT.
- Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Spina, Stefania (2001). *Fare i conti con le parole. Introduzione alla linguistica dei corpora*. Perugia: Guerra Edizioni.
- Stefanowitsch, Anatol (2020). *Corpus linguistics*. Berlin: Language Science Press.
- Steyer, Kathrin (2008). Kollokationen in deutschen Wörterbüchern und in der deutschen Wörterbuchforschung. In: *Lexicographica* 24. Tübingen: Niemeyer, 185-207.
- Steyer, Kathrin (2013). *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen: Narr Verlag.
- Storrer, Angelika (2013). Sprachstil und Sprachvariation in sozialen Netzwerken. In: B. Frank-Job/A. Mehler/T. Sutter (Hrsg.): *Die Dynamik sozialer und sprachlicher Netzwerke*. Wiesbaden: Springer VS, 331-366.
- Storrer, Angelika (2014). Sprachverfall durch internetbasierte Kommunikation? Linguistische Erklärungsansätze – empirische Befunde. In: A. Witt/A. Plewnia (Hrsg.). *Sprachverfall? Dynamik –Wandel – Variation*. Berlin: de Gruyter, 171-196.
- Storrer, Angelika (2018). Interaktionsorientiertes Schreiben im Internet. In: A. Deppermann/S. Reineke (Hrsg.). *Sprache im kommunikativen, interaktiven und kulturellen Kontext*. Berlin/Boston: de Gruyter, 219-244.
- Storrer, Angelika/Herzberg, Laura (2022). Alles okay! Korpusgestützte Untersuchungen zum Internationalismus OKAY. In: M. Beißwenger/L. Lemnitzer/C. Müller-Spitzer (eds.): *Forschen in der Linguistik: eine Methodeneinführung für das Germanistik-Studium*. Paderborn: UTB, 37-59.
- Tesnière, Lucien (1959). *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Thüne, Eva-Maria (2019). *Gerettet. Berichte von Kindertransport und Auswanderung nach Großbritannien*. Berlin/Leipzig: Hentrich & Hentrich.
- Thurmair, Maria (1989). *Modalpartikeln und ihre Kombinationen*. Tübingen: Niemeyer.
- Tognini-Bonelli, Elena (2001). *Corpus Linguistics at Work*. Amsterdam: Benjamins.
- Trawiński, Beata/Kupietz, Marc (2021). Von monolingualen Korpora über Parallel- und Vergleichskorpora zum Europäischen Referenzkorpus EuReCo. In: H. Lobin/A. Witt/A. Wöllstein (Hrsg.): *Sprachpolitisch, grammatisch, methodisch*.

- Jahrbuch des Instituts für Deutsche Sprache 2020. Berlin/Boston: de Gruyter, 209-234.
- Wallner, Franziska (2013). Korpora im Daf-Unterricht? Potentiale und Perspektiven am beispiel des DWDS. In: *Revista Nebrija de Lingüística Aplicada* (2013) 13. [https://www.nebrija.com/revista-linguistica/files/articulosPDF/articulo\\_533062d21f956.pdf](https://www.nebrija.com/revista-linguistica/files/articulosPDF/articulo_533062d21f956.pdf)
- Warren, Beatrice (1990). The importance of combining forms. In: W.U. Dressler (Hg.): *Contemporary morphology*. Berlin/New York: de Gruyter, 111-132.
- Weinrich, Harald (2005). *Textgrammatik der deutschen Sprache*. Hildesheim/Zürich/New York: Olms.
- Zifonun, Gisela/Hoffmann, Ludger/Strecker, Bruno (Hrsg.) (1997). *Grammatik der deutschen Sprache*. 3 Bände. Berlin/New York: de Gruyter.
- Zipf, George K. (1935). *The Psycho-biology of Language*. Boston: Hought Mifflin Co.
- Zipf, George K. (1949). *Human behaviour and the principle of least effort: an introduction to human ecology*. Cambridge: Addison-Wesley Press.
- Zimmermann, Malte (2011). Discourse Particles. In: K. von Heusinger/C. Maienborn/P. Portner (eds.): *Semantics*. Berlin: de Gruyter, 2011-2038.

### **Korpora, Tools und Ressourcen:**

- AntConc. [www.laurenceanthony.net/software/antconc/](http://www.laurenceanthony.net/software/antconc/)
- BNC – *British National Corpus*. [www.natcorp.ox.ac.uk](http://www.natcorp.ox.ac.uk)
- BW – *Berliner Wendekorpus*. PID = <http://hdl.handle.net/10932/00-0332-BD7C-3EF5-0B01-4>
- CORIS/CODIS. [https://corpora.ficlit.unibo.it/coris\\_ita.html](https://corpora.ficlit.unibo.it/coris_ita.html)
- CorpusExplorer – <https://notes.jan-oliver-ruediger.de/software/corpusexplorer-overview/>
- DeReKo – *Das Deutsche Referenzkorpus* <https://www.ids-mannheim.de/digspra/kl/projekte/korpora/>
- DGD – *Datenbank für Gesprochenes Deutsch*. [https://dgd.ids-mannheim.de/dgd/pragdb.dgd\\_extern.welcome](https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome)
- Duden 2016 = Wöllstein, Angelika/Dudenredaktion (Hrsg.). 2016. Dudenband 4 – *Die Grammatik*, 9. Aufl. Berlin: Dudenverlag.
- Duden 2022 = Wöllstein, Angelika/Dudenredaktion (Hrsg.). 2022. Dudenband 4 – *Die Grammatik*, 10. Aufl. Berlin: Dudenverlag.
- Duden – Wörterbuch. <https://www.duden.de>
- DWDS – *Digitales Wörterbuch der deutschen Sprache. Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart*. Hrsg. v. d. Berlin-Brandenburgischen Akademie der Wissenschaften. <https://www.dwds.de>

- EUROPARL – *European Parliament Proceedings Parallel Corpus*. <https://www.statmt.org/europarl/>
- E-Valbu = *Wörterbuch zur Verbalenz*; Grammatisches Informationssystem grammis, Leibniz-Institut für Deutsche Sprache, <https://grammis.ids-mannheim.de/verbalenz>
- Falko. <https://hu-berlin.de/falko>
- FOLK – *Forschungs- und Lehrkorpus Gesprochenes Deutsch*. PID = <http://hdl.handle.net/10932/00-0332-C1B2-A5E3-2A01-D>
- FOLKER. <https://agd.ids-mannheim.de/folker.shtml>
- f4x – Audiotranskription. <https://www.audiotranskription.de>
- Grammis = Leibniz-Institut für Deutsche Sprache: *Systematische Grammatik*. Grammatisches Informationssystem grammis. DOI:10.14618/grammatiksystem. <https://grammis.ids-mannheim.de/systematische-grammatik>
- GWSS – *Gesprochene Wissenschaftssprache Kontrastiv*. PID = <http://hdl.handle.net/10932/00-03BC-7412-E7EA-4101-3>
- InterCorp. <https://intercorp.korpus.cz/?lang=de>
- IS – *Emigrantendeutsch in Israel*. DGD, Leibniz-Institut für Deutsche Sprache. PID = <http://hdl.handle.net/10932/00-0332-C3A7-393A-8A01-3>
- ISW – *Emigrantendeutsch in Israel: Wiener in Jerusalem*. DGD, Leibniz-Institut für Deutsche Sprache. PID = <http://hdl.handle.net/10932/00-0332-C42A-423C-2401-D>
- ISZ – *Zweite Generation deutschsprachiger Migranten in Israel*. DGD, Leibniz-Institut für Deutsche Sprache. PID = <http://hdl.handle.net/10932/00-0332-C453-CEDC-B601-2>
- LBC – *Lessico dei Beni Culturali*. <http://corpora.lessicobeniculturali.net/de>
- MERLIN – *Multilingual Platform for European Reference Levels: Interlanguage Exploration in Context*. <https://merlin-platform.eu/index.php>
- OPUS – *Open Parallel Corpus*. <https://opus.nlpl.eu>
- Sketch Engine. <https://app.sketchengine.eu/>
- Virtual Language Observatory. <https://vlo.clarin.eu/?1>
- Word Smith Tools. <https://www.lexically.net/wordsmith/>
- Wortschatz Leibniz. <https://wortschatz.uni-leipzig.de/de>
- Zotero. <https://www.zotero.org/groups/2219390/israelkorpus/library>
- ZuMult – *Zugänge zu multimodalen Korpora gesprochener Sprache – Vernetzung und zielgruppenspezifische Ausdifferenzierung*. <https://zumult.org>

# Linguistica dei corpora

Una nuova prospettiva teorico-metodologica  
per lo studio di morfologia e sintassi in ambito *DaF*

Carolina Flinz

La linguistica dei corpora, la linguistica tedesca e la didattica del tedesco come lingua straniera (*Deutsch als Fremdsprache – DaF*) sono discipline strettamente collegate (cfr. Fandrych/Tschirner 2007). Nonostante questo, la loro sinergia per fini descrittivi, applicativi e didattici è ancora poco diffusa. Nella teoria e prassi didattica universitaria le connessioni sono infatti soprattutto di tipo binario, tra linguistica dei corpora e linguistica tedesca, tra linguistica tedesca e *DaF* e tra linguistica dei corpora e *DaF*. Obiettivo del presente volume è quello invece di far emergere le interconnessioni tra tutte e tre le discipline; svolgerà un ruolo centrale l'introduzione ad approcci di tipo empirico volti a favorire lo sviluppo della consapevolezza linguistica e crosslinguistica. Il focus del volume è su morfologia e sintassi, due ambiti linguistici tradizionalmente considerati centrali per la grammatica e che si prestano particolarmente a essere compresi e analizzati mediante la linguistica dei corpora. La loro analisi potrà servire da modello per altri livelli linguistici.

In copertina: immagine realizzata dall'autrice.

ISBN 979-12-55101-11-6 (print)  
ISBN 979-12-55101-13-0 (PDF)  
ISBN 979-12-55101-15-4 (EPUB)  
DOI 10.54103/milanoup.165