

# SWIKO ZIELHYPOTHESEN

## 1) GEGENSTAND (ZITIERT AUS FALKO HANDBUCH S.38FF)

Eine wichtige Komponente eines Lernerkorpus ist die Fehlerannotation. Um Fehler annotieren zu können, muss man sie identifizieren. Laut Corder (1986:37) müssen dafür die Lerneräußerungen mit den „korrekt“ rekonstruierten Äußerungen verglichen werden.

*We identify errors by comparing original utterances with what I shall call reconstructed utterances, that is, correct utterances having the meaning intended by the learner.*

Was man unter der rekonstruierten Äußerung zu verstehen hat, wird erkennbar, wenn man sich Lenbons (1991:182) Definition für Fehler ansieht. Unter einem Lernerfehler versteht er „a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speakers' native speaker counterparts.“ Dabei führt der Begriff der “rekonstruierten Äußerung” leicht in die Irre, denn eine direkte Verbindung zwischen Lerneräußerung und muttersprachlicher Rekonstruktion existiert nicht. Jede Lerneräußerung erlaubt eine Vielzahl gleichwertiger Äußerungen in der Muttersprache. Lüdeling (2008) zeigt in einer Studie mit verschiedenen professionellen Deutschlehrern, dass die Abweichung zwischen „Fehlerannotatoren“ sehr groß ist. Aus diesem Grund spricht sie von einer „Zielhypothese“, welche eine Interpretation der Lerneräußerung durch einen geschulten Annotator darstellt. Drei Konsequenzen erwachsen daraus. Zum Ersten muss eine Zielhypothese explizit sein, damit darauf basierende Fehlerannotationen transparent werden. Zweitens muss für Benutzer des Korpus die Möglichkeit bestehen, eigene, konkurrierende Zielhypothesen zu entwickeln und einzubinden. Drittens wird deutlich, dass die Erarbeitung einer Zielhypothese einer Operationalisierung mit mehr oder weniger engen Grenzen der Interpretation bedarf. Diese Richtlinien liegen hiermit vor. (...)

Eine **minimale Zielhypothese** dient als Normalisierungsebene und hat zum Ziel eine Ebene zu erzeugen, die für die automatische Verarbeitung dienen zu können. Die Anpassung des Originaltextes an eine parsbare Struktur ist somit höchste Priorität bei der Erstellung der minimalen Zielhypothese. Gleichzeitig wird versucht, die Anzahl der Abweichungen von der Originaläußerungen zu minimieren und nimmt dafür in Kauf, sich vor allem auf Fehler niedriger sprachlicher Ebenen (Orthografie, Morphologie, Syntax) zu beschränken.

## 2) AUSGANGSDATEI

Die Originaldatei besteht aus mehreren Transkriptions- und Annotationsebenen. Von besonderem Interesse für die vorliegenden Anweisungen sind die erste sowie zweitletzte Spur: Die *tok*-Spur enthält den ursprünglichen Lernertext, welcher nun in der *TH1*-Spur als 1. Zielhypothese orthographisch und grammatisch korrekt umformuliert wird. Auf den *lemma*, *cpos* sowie *SEA* werden ggf. nötige Informationen vermerkt.

	0	1	2	3	4	5	6
<b>Ex000 [tok]</b>	NA	Das	ist		1.	Beispiel	.
<b>Ex000 [ctok]</b>	NA	Das	ist		NUM	Beispiel	.
<b>Ex000 [lemma]</b>	NA	die	sein		@card@	Beispiel	.
<b>Ex000 [clemma]</b>							
<b>Ex000 [commonPOS]</b>	NA	DET	VER:PRE		NUM	NN	PUN
<b>Ex000 [lg-specific POS]</b>	NA	ART	VVFIN	ART	CARD	NN	\$.
<b>Ex000 [cpos]</b>							
<b>Ex000 [tag]</b>						error	
<b>Ex000 [markup]</b>					NUM		
<b>Ex000 [comment]</b>	ex.1						
<b>Ex000 [TH1]</b>							
<b>Ex000 [SEA]</b>							

### Originaltext des Lernenden in Token

Orthographisch korrigierter Originaltext

automatisch zugeteiltes Lemma

### manuell korrigiertes Lemma

Wortart (über alle 3 Sprachen)

Wortart (Deutsch spezifisch)

### manuell korrigierte Wortart

Fehler und nicht-Zielsprachliches

weitere Tags (z.B. Zahlen, Abkürzungen)

Kommentare des Transkribierenden

### 1. Zielhypothese

### semantische Fehlerannotation

### 3) VORGEHEN

Grundsätzlich gibt es 5 verschiedene Arten von Änderungen, welche im Nachfolgenden detailliert erklärt werden:

- i. **Verändern (CHA):** Ein (oder mehrere) Buchstabe(n) oder Wort muss verändert werden.
- ii. **Hinzufügen (INS):** ein Buchstabe oder Wort fehlt und muss hinzugefügt werden.
- iii. **Löschen (DEL):** Ein Buchstabe oder Wort ist überflüssig und muss entfernt werden.
- iv. **Verschieben (MOVS-MOVT):** Die Satzstellung muss angepasst werden, d.h. ein Wort oder mehrere Wörter muss / müssen verschoben werden (betrifft nur die Grammatik).
- v. **Verschmelzen oder trennen:** Zwei Wörter wurden fälschlicherweise getrennt (d.h. müssen verschmolzen werden: MERGE) respektive fälschlicherweise zusammengeschrieben (d.h. müssen getrennt werden: SPLIT).

Hier ein grammatisches Beispiel:

	0	1	2	3	4	5	6	7	8	9	10
Ex000 [tok]	In		Schweiz		es	gibt	8,2	Million	die	Tiere	.
Ex000 [TH1]	In	der	Schweiz	gibt	es		8,2	Millionen		Tiere	.
Ex000 [TH1_G_Diff]		INS		MOV-T		MOV-S		CHA	DEL		

Und hier ein orthographisches Beispiel:

	0	1	2	3	4	5	6	7	8	9
EXxxx [tok]	In	der	Schweiz	gibt	es	8,2	Milionen	Haus	Tiere	.
EXxxx [TH1]	In	der	Schweiz	gibt	es	8,2	Millionen	Haustiere		.
EXxxx [TH1_O_Diff]	SPLIT		DEL	CHA			INS	MERGE		

#### I. VERÄNDERN (CHA): EIN ODER MEHRERE BUCHSTABE/N ODER WORT MUSS VERÄNDERT WERDEN.

Wenn ein Wort falsch geschrieben wurde (z.B. Flexion), wird es in der TH1-Spalte korrigiert.

	0	1	2	3	4	5	6	7	8	9	10
Ex000 [tok]	In		Schweiz		es	gibt	8,2	Million	die	Tiere	.
Ex000 [ctok]	In		Schweiz		es	gibt	8,2	Million	die	Tiere	.
Ex000 [lemma]	in		Schweiz		es	geben	@card@	Million	die	Tier	.
Ex000 [clemma]				geben							
Ex000 [commonPOS]	PRP		NP		PRO:PER	VER:PRE	NUM	NN	DET	NN	PUN
Ex000 [lg-specific POS]	APPR		NE		PPER	VVFIN	CARD	NN	ART	NN	\$.
Ex000 [cpos]		ART									
Ex000 [tag]			error								
Ex000 [markup]											
Ex000 [comment]											
Ex000 [TH1]	In	der	Schweiz	gibt	es		8,2	Millionen		Tiere	.
Ex000 [SEA]											

ACHTUNG: Handelt es sich dabei um ein falsches Inhaltswort in derselben Wortart (z.B. *haben* statt *sein*) , wird zusätzlich das korrigierte Lemma in der *clemma*-Spur eingetragen.

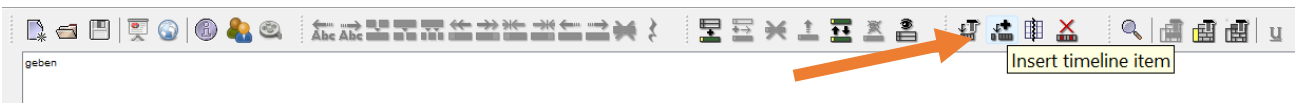
	0	1	2	3	4	5
Ex000 [tok]	Ich	habe	15	Jahre	alt	.
Ex000 [ctok]	Ich	habe	15	Jahre	alt	.
Ex000 [lemma]	ich	haben	@card@	Jahr	alt	.
Ex000 [clemma]		sein				
Ex000 [commonPOS]	PRO:PER	VER:PRE	NUM	NN	ADJ	PUN
Ex000 [lg-specific POS]	PPER	VAFIN	CARD	NN	ADJD	.\$
Ex000 [cpos]						
Ex000 [tag]						
Ex000 [markup]						
Ex000 [comment]						
Ex000 [TH1]	Ich	bin	15	Jahre	alt	.
Ex000 [SEA]						

ACHUNGTG: Handelt es sich dabei um eine falsches Wort in einer anderen Wortart (z.B. *warum* statt *weil*), wird zusätzlich die korrigierte Wortart in die *cpos*-Spur eingetragen. Dies gilt auch, wenn ein Komma mit *und* ersetzt wird.  
 ACHTUNG: Wird ein anderes Wort als in der bereits korrigierten *ctok*-Spur gewählt, weil es auf ein Wort aus einer anderen Sprache zurückgeführt werden kann (z.B. *Zelt* statt *tente*), muss der entsprechende Sprachenverweis auf der *markup*-Spur eingetragen werden (*schlg* für Schulsprache, *foreignEng* für Englisch).

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Ex000 [tok]	Das	camping	ist	nicht	gut		warum	ich	mag	niicht	shclafen	im	eine	tante			.
Ex000 [ctok]	Das	Camping	ist	nicht	gut		warum	ich	mag	nicht	schlafen	im	eine	Tante			.
Ex000 [lemma]	die	Camping	sein	nicht	gut		warum	ich	mögen	nicht	schlafen	in+die	eine	Tante			.
Ex000 [clemma]															schlafen	mögen	
Ex000 [commonPOS]	DET	NN	VER:PRE	PAR	ADJ		PRO	PRO:PER	VER:PRE	PAR	VER:PRE	PRP:DET	DET	NN			
Ex000 [lg-specific POS]	ART	NN	VAFIN	PTKNEG	ADJD		PWAV	PPER	VMFIN	PTKNEG	VVFIN	APPRART	ART	NN			
Ex000 [cpos]							KOUS										
Ex000 [tag]		error								error	error			error			
Ex000 [markup]														schlg_item			
Ex000 [comment]																	
Ex000 [TH1]	Das	Camping	ist	nicht	gut	,	weil	ich		nicht		in	einen	Zelt	schlafen	mag	.
Ex000 [SEA]																	

# I. HINZUFÜGEN (INS): EIN BUCHSTABE ODER WORT FEHLT UND MUSS HINZUGEFÜGT WERDEN.

Um ein fehlendes Wort oder Satzzeichen hinzuzufügen, muss eine neue Spalte hinzugefügt werden:



	0	1	2	3	4	5	6	7	8	9	10	11	12
Ex000 [tok]	In	Schweiz		es	gibt	8,2	Million	die	Tiere	.			
Ex000 [ctok]	In	Schweiz		es	gibt	8,2	Million	die	Tiere	.			
Ex000 [lemma]	in	Schweiz		es	geben	@card@	Million	die	Tier	.			
Ex000 [clemma]			geben										
Ex000 [commonPOS]	PRP	NP		PRO:PER	VER:PRE	NUM	NN	DET	NN	PUN			
Ex000 [lg-specific POS]	APPR	NE		PPER	VVFIN	CARD	NN	ART	NN	\$.			
Ex000 [cpos]													
Ex000 [tag]		error											
Ex000 [markup]													
Ex000 [comment]													
Ex000 [TH1]	In	Schweiz	gibt	es		8,2	Millionen		Tiere	.			
Ex000 [SEA]													

Anschliessend müssen zwei Informationen ergänzt werden:

- 1) das neue Wort in der TH1-Spalte; und
- 2) die entsprechende Wortart («POS-Tagset Deutsch», welches der lg-specific POS-Spur entspricht) in der cpos-Spalte.

	0	1	2	3	4	5	6	7	8	9	10
Ex000 [tok]	In	Schweiz		es	gibt	8,2	Million	die	Tiere	.	
Ex000 [ctok]	In	Schweiz		es	gibt	8,2	Million	die	Tiere	.	
Ex000 [lemma]	in	Schweiz		es	geben	@card@	Million	die	Tier	.	
Ex000 [clemma]			geben								
Ex000 [commonPOS]	PRP	NP		PRO:PER	VER:PRE	NUM	NN	DET	NN	PUN	
Ex000 [lg-specific POS]	APPR	NE		PPER	VVFIN	CARD	NN	ART	NN	\$.	
Ex000 [cpos]		ART									
Ex000 [tag]		error									
Ex000 [markup]											
Ex000 [comment]											
Ex000 [TH1]	In	der	Schweiz	gibt	es		8,2	Millionen		Tiere	.
Ex000 [SEA]											

## II. LÖSCHEN (DEL): EIN BUCHSTABE ODER WORT IST ÜBERFLÜSSIG UND MUSS ENTFERNT WERDEN.

Ist ein Wort überflüssig, wird es aus der TH1-Spalte gelöscht.

ACHTUNG: Die Zelle muss anschliessend komplett leer sein, d.h. nicht mehr weiss sondern grau hinterlegt.

	0	1	2	3	4	5	6	7	8	9	10
Ex000 [tok]	In		Schweiz		es	gibt	8,2	Million	die	Tiere	.
Ex000 [ctok]	In		Schweiz		es	gibt	8,2	Million	die	Tiere	.
Ex000 [lemma]	in		Schweiz		es	geben	@card@	Million	die	Tier	.
Ex000 [clemma]				geben							
Ex000 [commonPOS]	PRP		NP		PRO:PER	VER:PRE	NUM	NN	DET	NN	PUN
Ex000 [lg-specific POS]	APPR		NE		PPER	VVFIN	CARD	NN	ART	NN	\$.
Ex000 [cpos]		ART									
Ex000 [tag]			error								
Ex000 [markup]											
Ex000 [comment]											
Ex000 [TH1]	In	der	Schweiz	gibt	es		8,2	Millionen		Tiere	.
Ex000 [SEA]											

IV. VERSCHIEBEN (MOVS-MOVT) (NUR GRAMMATIK): DIE SATZSTELLUNG MUSS ANGEPA SST WERDEN, D.H. EIN WORT ODER MEHRERE WÖRTER MUSS / MÜSSEN VERSCHOBEN WERDEN.

Fehler in der Satzstellung müssen korrigiert werden. Dies benötigt folgende Schritte:

- 1) Eine neue Spalte wird am entsprechenden Ort hinzugefügt (siehe 1. hinzufügen);
- 2) Auf der TH1-Spur wird der Satz in der korrekten Satzstellung notiert;
- 3) Auf der *lemma*-Spur wird das Lemma des verschobenen Wortes vermerkt.

	0	1	2	3	4	5	6	7	8	9	10
Ex000 [tok]	In		Schweiz		es	gibt	8,2	Million	die	Tiere	.
Ex000 [ctok]	In		Schweiz		es	gibt	8,2	Million	die	Tiere	.
Ex000 [lemma]	in		Schweiz		es	geben	@card@	Million	die	Tier	.
Ex000 [clemma]				geben							
Ex000 [commonPOS]	PRP		NP		PRO:PER	VER:PRE	NUM	NN	DET	NN	PUN
Ex000 [lg-specific POS]	APPR		NE		PPER	VVFIN	CARD	NN	ART	NN	\$.
Ex000 [cpos]		ART									
Ex000 [tag]			error								
Ex000 [markup]											
Ex000 [comment]											
Ex000 [TH1]	In	der	Schweiz	gibt	es		8,2	Millionen		Tiere	.
Ex000 [SEA]											

Detaillierte Regeln dazu, welches Wort oder welche Wörter in welche Richtung verschoben werden, folgen im nächsten Kapitel. «

SPEZIALFALL VERSCHIEBEN + ÄNDERN:

Das neu eingefügte *lemma* muss genau so im Originaltext gefunden werden können. Falls nötig, muss beim ursprünglichen Lemma ebenfalls ein *lemma* eingefügt werden (vgl. unter Punkt III verändern).

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Ex000 [tok]		die	Ferien	auf	die	Berg	,		wir	haben	skifahren	mit	mein	Onkle	und	meine	Tante				.	
Ex000 [ctok]		die	Ferien	auf	die	Berg	,		wir	haben	Skifahren	mit	mein	Onkel	und	mein	Tante				.	
Ex000 [lemma]		Die	Ferien	auf	die	Berg	,		wir	haben	Skifahren	mit	mein	Onkel	und	meine	Tante				.	
Ex000 [clemma]						sein			sein	Ski	fahren									Ski	fahren	.
Ex000 [commonPOS]		DET	NN	PPR	ART	NN	\$,		PRO:PER	VER:PRE	NN	PRP	PRO:POS	NN	KON	POS	NN				PUN	
Ex000 [lg-specific POS]		ART	NN	APPR	ART	NN	\$,		PPER	VAFIN	NN	APPR	PPOSAT	NN	KON	PPOSAT	NN				\$.	
Ex000 [cpos]	APPR																					
Ex000 [tag]		error									error				error							
Ex000 [markup]																						
Ex000 [comment]																						
Ex000 [TH1]	In	den	Ferien	auf	dem	Berg		sind	wir				mit	meinem	Onkel	und	meiner	Tante		Ski	gefahren	
Ex000 [SEA]																						

V. VERSCHMELZEN ODER TRENNEN: ZWEI WÖRTER WURDEN FÄLSCHLICHERWEISE GETRENNT (D.H. MÜSSEN VERSCHMOLZEN WERDEN: MERGE) RESPEKTIVE FÄLSCHLICHERWEISE ZUSAMMENGESCHRIEBEN (D.H. MÜSSEN GETRENNT WERDEN: SPLIT).

Fälschlicherweise getrennt geschriebene Wörter werden korrigiert und die Zellen via *merge* verschmolzen. Falls nötig wird ein vorhandenes «NA» dabei gelöscht, so dass nur noch das Wort in der Zelle übrig bleibt.

Fälschlicherweise zusammengeschriebene Wörter werden in der TH1-Spur getrennt geschrieben. Dabei muss darauf geachtet werden, dass die Zelle, in der das Wort gelöscht wurde, komplett leer ist (d.h. nicht mehr weiss, sondern grau hinterlegt).

	2	3	4
Ex000 [tok]	die	forêt	noir
Ex000 [ctok]	di	Schwarzwäldertorte	NA
Ex000 [lemma]	die	Schwarzwäldertorte	NA
Ex000 [clemma]			
Ex000 [commonPOS]	DET	NN	NA
Ex000 [lg-specific POS]	ART	NN	NA
Ex000 [cpos]			
Ex000 [tag]		schlg_item	
Ex000 [markup]			
Ex000 [comment]			
Ex000 [TH1]	die	Schwarzwäldertorte	
Ex000 [SEA]			

	0	1	2	3
Ex000 [tok]	In	der	Schule	
Ex000 [ctok]	In	der	Schule	
Ex000 [lemma]	in	die	Schule	
Ex000 [clemma]				
Ex000 [commonPOS]	PRP	DET	NN	
Ex000 [lg-specific POS]	APPR	ART	NN	
Ex000 [cpos]				
Ex000 [tag]	error	error		
Ex000 [markup]				
Ex000 [comment]				
Ex000 [TH1]	In	der	Schule	
Ex000 [SEA]				

#### 4) RICHTLINIEN

*ORTHOGRAPHISCH, MORPHOLOGISCH UND SYNTAKTISCH KORREKT  
SO WENIG WIE MÖGLICH, SO VIEL WIE NUR ABSOLUT NÖTIG*

1. **GERINGE KORREKTUREN:** Was nicht geändert werden muss, wird nicht verändert. Je weniger Tokens verändert werden müssen, desto besser. Tokens werden so in einem Satz verschoben, dass möglichst wenige Konstituenten und Tokens davon betroffen sind.

ctok	Frauen	konnten		solchen		gesellschaftlichen	Zustand	verändern
ZH1	Frauen	konnten	einen	solchen		gesellschaftlichen	Zustand	verändern
-ZH1	Frauen	konnten		solch	einen	gesellschaftlichen	Zustand	verändern

(fk024\_2006\_07)

tok	Ich	mag					gern	Ferien	mit	der	Familie		.
TH1	Ich	mag						Ferien	mit	der	Familie	gern	.
-TH1	Ich	mag	Ferien	mit	der	Familie	gern						.

2. **BEWEGEN STATT TAUSCHEN:** Zielposition für eine bewegte Konstituente in der Zielhypothese ist immer ein leeres Token. Konstituenten werden nicht getauscht.

tok	In	der	Schweiz		es	gibt	8,2	Million	Tiere	.
TH1	In	der	Schweiz	gibt	es		8,2	Millionen	Tiere	.
-TH1	In	der	Schweiz		gibt	es	8,2	Millionen	Tiere	.

3. **BEWEGUNG NACH LINKS:** Spricht keine der vorangegangenen Regeln dagegen, werden Bewegungen nach links denen nach rechts vorgezogen.

ctok		Die	Frauen	haben		die	Macht	auch	.
ZH1		Die	Frauen	haben	auch	die	Macht		.
-ZH1	Auch	die	Frauen	haben		die	Macht		.

4. **KURZE BEWEGUNGEN:** Weicht die Position eines Tokens in der Zielhypothese von ihrem Pendant auf der tok-Ebene ab, so soll der Abstand zwischen beiden Positionen minimal bleiben.

Tok	In	der	Schweiz		es	gibt		8,2	Millionen	Tiere	.
TH1	In	der	Schweiz	gibt	es			8,2	Millionen	Tiere	.
-T1	In	der	Schweiz			gibt	es	8,2	Millionen	Tiere	.

5. **OPERATIONEN:**

> VERSCHMELZEN, TRENNEN UND DOPPELUNGEN LÖSCHEN: Zellen werden ebenfalls verschmolzen oder aufgeteilt respektive rechts gelöscht.

> VERÄNDERN > HINZUFÜGEN > LÖSCHEN: Spricht keine der vorangegangenen Regeln dagegen, wird wenn möglich verändert, bevor Token hinzugefügt respektive gelöscht werden.

Im Folgenden wird auf die einzelnen Regeln ausführlicher eingegangen.



## GRUNDSATZ 1: GERINGE KORREKTUREN

Was nicht geändert werden muss, wird nicht verändert. Je weniger Tokens verändert werden müssen, desto besser. Tokens werden so in einem Satz verschoben, dass möglichst wenige Konstituenten und Tokens davon betroffen sind.

### SYNTAX

Verbstellung in Haupt- und Nebensätzen wird korrigiert. (SWIKO anders, MERLIN/FALKO linke Satzklammer beibehalten)

Die Verbstellung wird sowohl in Haupt- als auch Nebensätzen korrigiert. Das finite Verb darf sowohl in der linken als auch rechten Satzklammer bewegt werden.

tok	In		Schweiz		es	gibt		8,2	Millionen	Tiere	.
TH1	In	der	Schweiz	gibt	es			8,2	Millionen	Tiere	.
-TH1	In	der	Schweiz			gibt	es	8,2	Millionen	Tiere	.

In subordinierten Nebensätzen wird Verbzweitstellung korrigiert. Findet sich in einem Nebensatz eine falsche Kombination aus Konnektorvorgabe und Wortstellung, so wird die Wortstellung auch dann angepasst, wenn eine Veränderung des Konnektors weniger Abweichung insgesamt bedeuten würde.

tok	,	weil	ich	habe	zwei	Hund	in	mein	house		.
TH1	,	weil	ich		zwei	Hunde	in	meinem	Haus	habe	.
tok	,	denn	ich	habe	zwei	Hunde	in	meinem	Haus		.

Ein fehlendes oder redundantes «zu» in Infinitivkonstruktionen wird korrigiert, d.h. in TH1 hinzugefügt oder weggelassen, da es sich um einen Fehler bei der Verwendung grammatikalischer Ergänzungen handelt. (Merlin)

tok	Gehen	in	eine	oder	countri			das	ist	nicht	gut	.
TH1		In	ein	anderes	Land	zu	gehen		ist	nicht	gut	.

Abfolgen von Konstituenten im Mittelfeld werden nur dann korrigiert, wenn dadurch ungrammatische Strukturen entstehen. Wortstellungstendenzen werden ignoriert.

tok	Es	ist			nicht	gut	für	Jugendliche
TH1	Es	ist			nicht	gut	für	Jugendliche
-TH1	Es	ist	für	Jugendliche	nicht	gut		

Tauchen in einem Satz der «dass»-Subordinierer und ein Infinitiv mit «zu» auf, wird die Variante mit der kleineren Abweichung gewählt.

*ctok: Aber... Einige streben danach, dass noch reicher zu sein, noch gewaltiger; (usb015\_2006\_10)*

*TH1: Aber: Einige streben danach, noch reicher zu sein und noch gewaltiger;*

### FEHLENDE SATZGLIEDER

Fehlende Funktionswörter werden passend zum Kontext ergänzt. (Merlin)

tok	In		Schweiz
TH1	In	der	Schweiz

Fehlende Inhaltswörter werden nur ergänzt, wenn sie erforderlich sind, um einen syntaktisch wohlgeformten Satz sicherzustellen.

Elliptische Sätze sind ausgeschlossen. In der SEA-Spur (*semantic error annotation*) wird festgehalten, ob die Ergänzung zwar unsicher aber aus dem Kontext ableitbar (1) oder eine wilde Vermutung (2) ist. Falls nötig wird sowohl Subjekt als auch Prädikat ergänzt. (nur MERLIN, FALKO korrigiert weder Verberstdeklarativsätze ohne Subjekt, noch verblöse oder uneingeleitete oder isoliert stehende Nebensätze).

tok	Die	Kleidung		im		50eme	placs	.
TH1	Die	Kleidung	ist	auf	dem	50.	Platz	.
SEA			1					

tok	Ich	mag			nicht	winter		weil	er		ist	.
TH1	Ich	mag	den	Winter	nicht		,	weil	er	kalt	ist	.
SEA										2		

tok			8,2	millionen	animaux	Domestiquieren	.
TH1	Es	gibt	8,2	Millionen	Haustiere		.
SEA	2	2					

Beispiel Ellipse: Mein Bruder heisst Pascal und meine Schwester Laura.

SWIKO-Beispiel Ellipse: Jeder Jugendliche nutzt es und mittlerweile auch mehr ältere Menschen.

Ausnahme: «Titel» bei Präsentationen, z.B. «Die Sprachen.»

## SATZZEICHEN

Wurde ein Hauptsatz von einem zweiten, koordinierten Hauptsatz mit Komma getrennt, wird das Komma mit einem Punkt ersetzt.

ctok: Hallo, ich heisse Laura und ich komme aus der Schweiz, ich habe 15 Jahre Alt.

TH1: Hallo, ich heisse Laura und ich komme aus der Schweiz. Ich bin 15 Jahre alt.

Zwei oder mehr unvollständige, mit einem Punkt getrennte Satzteile werden mit einem Komma verbunden.

ctok: Es gibt 8,2 millionen animaux domestiquieren . 8,2 millionen fisch, 1,5 millionen katz, 1/2 millionen hund.

ZH1: Es gibt 8,2 Millionen Haustiere, 8,2 Millionen Fische , 1,5 Millionen Katzen, ½ Million Hunde.

AUSNAHME: Wurden die Satzteile vom Lernenden durch einen Paragraphen getrennt (auf der markup-Spur mit «par» markiert), dürfen sie nicht verbunden werden. Stattdessen werden zwei Sätze daraus formuliert.

tok	Ja	,	ich	liebe	telefonieren	und	fliegen	.			fernsehen	und	jacke	.	
markup									par						
TH1	Ja	,	ich	liebe	telefonieren	und	fliegen	.		Ich	ergänze	fernsehen	und	Jacken	.
SEA										2	2				
-TH1	Ja	,	ich	liebe	telefonieren	und	fliegen	,				fernsehen	und	Jacke	.

Zweiteilige Konnektoren müssen in einem Matrixsatz verbunden werden.

Beispiele: zwar ..., aber / je ... desto / nicht nur ..., sondern auch

ctok: Ich liebe Sprachen. Ich kann sie zwar meistens nicht so gut. Trotzdem möchte ich Übersetzerin werden.

ZH1: Ich liebe Sprachen. Ich kann sie zwar meistens nicht so gut, aber trotzdem möchte ich Übersetzerin werden.

Falsch gesetzte Kommas werden am richtigen Ort eingefügt (mit Vermerk auf der clemma-Spur).

tok	Das	Ist	good		weil	,	
TH1	Das	ist	gut	,	weil		
-TH1	Das	ist	gut			,	weil

## FLEXION

Subjekt-Verb-Agreement und Flexion werden angepasst, auch wenn das Hinzufügen von Tokens eine Alternative mit gleich vielen grammatischen Änderungen wäre. (SWIKO)

tok	mein	hobby	ist	fussball	spielen	und				boxe	.
TH1	Meine	Hobbies	sind	Fussball	spielen	und				boxen	.
-TH1	Mein	Hobby	ist	Fussball	spielen	und	mein	Hobby	ist	boxen	.

Wenn die semantische oder pragmatische Struktur einer Äußerung einen bestimmten Tempus oder Modus erfordert und dieser nicht eingehalten werden soll, erfolgt in TH1 keine Korrektur. ABER: Der Konflikt wird aber auf der SEA-Spur festgehalten mit «VF» (Verbformation) festgehalten (d.h. tense= past/present/future, mood = imperative/conjunctive/indicative, voice = active/passive) – aber nur, wenn das Verb nicht bereits korrigiert werden musste. (SWIKO Zusatz)

tok	Erst	verliert	er	das	Geld	und	dann	floh	er	.
TH1	Erst	verliert	er	das	Geld	und	dann	floh	er	.
-TH1	Erst	verlor	er	das	Geld	und	dann	floh	er	.
SEA		VF								

Verb als Kern: Stimmen das Verb und seine Argumente nicht überein, werden das Verb beibehalten und die Argumente angepasst. Für fehlende obligatorische Objekte und für vom Verb abhängige Präpositionalobjekte werden Dummys eingefügt. Numerus und Kasus richten sich nach dem Verb.

ctok	dass	unsere	Zivilisation	sich	aus	Gruppen	besteht	.
ZH1	dass	unsere	Zivilisation		aus	Gruppen	besteht	.
-ZH1	dass	unsere	Zivilisation	sich	aus	Gruppen	zusammensetzt	.

tok	Die	Brille	sind	sehr	wichtig
TH1	Die	Brillen	sind	sehr	wichtig
-TH1	Die	Brille	ist	sehr	wichtig

**AUSSER:** Können dem Verb keine Argumente hinzugefügt werden, wird das Verb präfigiert oder mit einem Partikel versehen.

*Ctok: ..., es sei denn der Staat **passt** neue Arbeitsmarkt regelnde Gesetze.*

*TH1: ..., es sei denn, der Staat **passt** neue arbeitsmarktregelnde Gesetze **an**.*

Nominale Kongruenzhierarchie: Bei fehlender Kongruenz zwischen Determinierer, Attribut und Nominalkern wird vom Kern ausgegangen. Danach wird zugunsten des Determinierers entschieden: Attribute < Determinierer (syntaktischer Kopf) < syntaktischer Kern (lexikalischer Kopf)

ctok	keine	Arbeit	bekommen	,	die	ihrer	Qualifikationen	entspricht	.
ZH1	keine	Arbeit	bekommen	,	die	ihren	Qualifikationen	entspricht	.
-ZH1	keine	Arbeit	bekommen	,	die	ihrer	Qualifikation	entspricht	.

Kasusrektion durch Präpositionen wird gemäss der Darstellung in Grammis (Breindl 2000) behandelt, hierbei werden auch progressive Entwicklungen hin zum Dativ akzeptiert.

z.B. «Während den Sechzigerjahren» wird nicht korrigiert.

## LEXIK

Werden ganze Wörter in Grossbuchstaben geschrieben, um etwas hervorzuheben, werden diese in Standard-Grossschreibung umgeschrieben. In der *markup* Zeile wird es mit *emphasis* vermerkt. (*Dies sollte bereits der Fall sein*).

tok	Ich	denke	die	Jugendherberge	sind	nicht	IN
TH1	Ich	denke	die	Jugendherberge	sind	nicht	in
markup							emphasis

Semantische und pragmatische Aspekte werden nicht korrigiert: Falsch verwendete Ausdrücke werden nur korrigiert, wenn sie zu einem syntaktisch falschen Satz oder grammatisch fehlerhaften Formen führen.

**ABER:** Bei unangemessen verwendeten Wörtern wird *WW* (Wortwahl) respektive *WF* (Wortbildung) in der *SEA*-Spur (semantic error annotation) vermerkt.

*Beispiel Nomen und Verben:*

tok	Eine	Hut	kostet	3500	Franken	.
TH1	Ein	Hut	kostet	3500	Franken	.
SEA		WW				.

tok	Für	einen	Hund	kaufen	wir	1500	für	das	Essen	.
TH1	Für	einen	Hund	kaufen	wir	1500	für	das	Essen	.
SEA				WW						.

*Beispiel Konnektoren:*

Wenn die Wahl des Konnektors zu einem syntaktisch falschen Satz führt, wird dieser korrigiert. Eine semantisch/pragmatisch falsche Wahl wird nur auf der *SEA*-Spur vermerkt.

tok	Ich	weiss	nicht	,	dass	du	kommst	.
TH1	Ich	weiss	nicht	,	dass	du	kommst	.
SEA					WW			.

*Beispiel Komparativ- / Superlativ-Formen:*

Grammatisch fehlerhafte Formen werden korrigiert. Die unangemessene Verwendung korrekter Komparativ-/Superlativformen, bei denen kein syntaktischer Konflikt auftritt, wird auf der *SEA*-Spur vermerkt.

tok	Aber	die	Elektrizität	ist	Besser	wichtig	.
TH1	Aber	die	Elektrizität	ist	wichtiger		.

*Beispiel Komposita:*

Wortbildungsfehler einschliesslich fehlender oder falscher «Fugenelemente» werden auf der *SEA*-Spur festgehalten.

tok	geschmack	idea
TH1	Geschmacksidee	

SEA	WB
-----	----

tok						.
TH1						.
SEA						.

Ggf. Beispiele zu lexematisch geblockten aber morphologisch zulässigen Varianten sowie fehlenden Mitteln zum korrekten Ausdruck von Sprechereinstellungen (werden nicht korrigiert, aber ggf. auf der SEA-Spur vermerkt?)

**Entities:** Eine Korrektur von Eigennamen in TH1 ist nur bei Toponymen und Namen von Personen/Institutionen zu erwarten, wenn die genannte Einheit eindeutig ist. Nicht-Zielsprachliches wird nur übersetzt, wenn es sich um Toponyme handelt; bei Eigennamen wie z.B. Bandnamen, Lied- oder Buchtitel wird nur Gross-/Kleinschreibung korrigiert, ansonsten die Originalform beibehalten.

tok	Meine	Lieblingsmusik	Ist	PNL	autre	monde	.
TH1	Meine	Lieblingsmusik	Ist	PNL	Autre	Monde	.

tok	Ich	habe	snapchat	und	tiktok	und	instagrame	.
TH1	Ich	habe	Snapchat	und	Tiktok	und	Instagram	.

tok	Mein	Lieblingsbuch	heisst	«	who	did	it	»
TH1	Mein	Lieblingsbuch	heisst	«	who	did	it	»

**Abkürzungen und Zahlen** sind entsprechend in der *markup*-Spur vermerkt. Grundsätzlich werden sie übernommen, wie sie in der Originalspur sind mit Ausnahme von geläufigen Abkürzungen (z.B., etc.), die korrigiert werden.

	44	45	46	47	48	49	50	51	52
RI207 [tok]	ich	möchte	eine	Motocross	,	eine	125cm3	yz	.
RI207 [ctok]	ich	möchte	eine	Motocross	,	eine	NUM	ABV	.
RI207 [lemma]	ich	mögen	eine	Motocross	,	eine	@card@	ABV	.
RI207 [clemma]									
RI207 [commonPOS]	PRO:PER	VER:PRE	DET	NN	\$,	DET	NUM	NP	PUN
RI207 [lg-specific POS]	PPER	VMFIN	ART	NN	\$,	ART	CARD	NE	\$.
RI207 [cpos]									
RI207 [tag]									
RI207 [markup]							NUM	ABV	
[comment]									
RI207 [TH1]	ich	möchte	eine	Motocross	,	eine	125cm3	yz	.
RI207 [SEA]									

Ausnahmen: 1 Haus von 2 (Ein Haus von zwei); Ich bin 14 Jahre alt und habe 1 Bruder und 1 Schwester (Ich bin 14 Jahre alt und habe einen Bruder und eine Schwester).

**Unlesbare Satz(teile)** wurden in der Transkription mit </unreadable> oder <unknown> gekennzeichnet. Falls eine Interpretation möglich ist, wird diese festgehalten mit entsprechendem «2» Vermerkt auf der SEA-Spur. Falls nicht, wird das Tag beibehalten.

	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Es152 [tok]	In		Schweiz				barcht	millions	zweihundert	Tiere	die	Leute	haben	.
Es152 [ctok]	In		Schweiz				NA	Millionen	zweihundert	Tiere	die	Leute	haben	.
Es152 [lemma]	in		Schweiz				NA	Million	zweihundert	Tier	die	Leute	haben	.
Es152 [clemma]				haben	die	Leute								
Es152 [commonPOS]	PRP		NP				NA	NN	NUM	NN	DET	NN	VER:INF	PUN
Es152 [lg-specific POS]	APPR		NE				NA	NN	CARD	NN	ART	NN	VAINF	\$.
Es152 [cpos]														
Es152 [tag]								schlg_item						
Es152 [markup]							unknown							period
[comment]														
Es152 [TH1]	In	der	Schweiz	haben	die	Leute	aacht	Millionen	zweihundert	Tiere	die	Leute	haben	.
Es152 [SEA]							1							

**Symbole** werden falls sinnvoll «übersetzt». Das *symbol*-Tag ist auf der markup-Spur vermerkt, das zu übersetzende Symbol auf der *comment*-Spur festgehalten.

	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
RI507 [tok]	Ein	Katze	isst	50.000	(	Reptiles	)	NA	250.000	(	papillons	)	NA	400	(	souris	)	pro	Monat	.
RI507 [ctok]	Ein	Katze	isst	NUM	(	Reptilien	)	NA	NUM	(	Schmetterlinge	)	NA	NUM	(	Mäuse	)	pro	Monat	.
RI507 [lemma]	eine	Katze	essen	@card@	(	Reptil	)	NA	@card@	(	Schmetterling	)	NA	@card@	(	Maus	)	pro	Monat	.
RI507 [clemma]																				
RI507 [commonPOS]	DET	NN	VER:PRE	NUM	PUN	NN	PUN	NA	NUM	PUN	NN	PUN	NA	NUM	PUN	NN	PUN	PRP	NN	PUN
RI507 [lg-specific POS]	ART	NN	VVFIN	CARD	\$(	NN	\$(	NA	CARD	\$(	NN	\$(	NA	CARD	\$(	NN	\$(	APPR	NN	\$.
RI507 [cpos]																				
RI507 [tag]						schlg_item					schlg_item						schlg_item			
RI507 [markup]				NUM				symbol	UM				symbol	UM						
[comment]								+					+							
RI507 [TH1]	Ein	Katze	isst	NUM	(	Reptilien	)	und	UM	(	Schmetterlinge	)	und	UM	(	Mäuse	)	pro	Monat	.
RI507 [SEA]																				

### Nicht-Zielsprachliches

Wörter aus anderen Sprachen werden wenn möglich übersetzt. Bei mehreren Möglichkeiten wird das Wort gewählt, das weniger weitere Änderungen verursacht, ausser wenn diese Alternative sehr selten und/oder stark umgangssprachlich ist (Beispiel b).

tok	Die	inventionen	sind	sehr	praktisch	.
TH1	Die	Erfindungen	Sind	sehr	praktisch	.

tok	Ich	ajoutire	der	Ferien	In	Italy	.	
TH1	Ich	ergänze		Ferien	in	Italien	.	
-TH1	Ich	füge		Ferien	in	Italien	hinzu	.

tok	Wir	dependieren	1000	frs	pro	jahr	.	
TH1	Wir	geben	1000	frs	pro	Jahr	aus	.
-TH1	Wir	verprassen	1000	frs	pro	Jahr		.

Ausnahme: Das betreffende Wort / der betreffende Satz wird zur Erklärung/Spezifizierung des L2-Äquivalents verwendet und daher in Klammern oder Anführungszeichen geschrieben. (nur Merlin)

*tok*: Im Franz lernen wir Sachen wie z.B. Die Heuschrecke la sauterell. Braucht man einfach nicht.

*TH1*: Im Französisch lernen wir Sachen wie z.B. die Heuschrecke la sauterelle. Das braucht man einfach nicht.

## GRUNDSATZ 2: BEWEGEN STATT TAUSCHEN

Zielposition für eine bewegte Konstituente in der Zielhypothese ist immer ein leeres Token. Konstituenten werden nicht getauscht.

## GRUNDSATZ 3: MINIMALE BEWEGUNGEN

Weicht die Position eines Tokens in der Zielhypothese von ihrem Pendant auf der tok-Ebene ab, so soll der Abstand zwischen beiden Positionen minimal bleiben.

### AUSNAHMEN

Führt eine minimale Bewegung zu einer Struktur, die einerseits sehr stark von der Originalstruktur abweicht, die intendierte Zielstruktur aber andererseits nicht gut abbildet, so darf die minimale Bewegung verletzt werden.

ctok		In	Neu	Seeland	auch	haben		wir	eine	Prime	Minister
ZH1	Auch	in	Neuseeland			haben		wir	eine	Preministerin	
-ZH1		In	Neuseeland			haben	auch	wir	eine	Preministerin	

Führt eine kurze Bewegung zur Teilung koordinierter Strukturen, wird eine längere Bewegung bevorzugt.

Ich bin dafür für ferien in land ist sehr langweilig und sehr long die tag.

→ Ich bin dafür, dass Ferien im Land sehr langweilig und die Tage sehr lang sind.

→ NICHT: Ich bin dafür, dass Ferien im Land sehr langweilig sind und die Tage sehr lang.

## GRUNDSATZ 4: BEWEGUNG NACH LINKS

Spricht keine der vorangegangenen Regeln dagegen, werden Bewegungen nach links denen nach rechts vorgezogen.

## GRUNDSATZ 5: OPERATIONEN

### TOKEN VERSCHMELZEN & AUFTEILEN

Werden zwei Token miteinander verschmolzen, werden auch die Zellen zu einer Spanne verschmolzen. Analog dazu, wird ein Token getrennt, wird auch die Zeile aufgeteilt.

ctok	an	dem	Wohlstand
ZH1	am		Wohlstand
~ZH1		am	
~ZH1	am		Wohlstand

ctok	zum		Frau
ZH1	zu	der	Frau

Wurden in einem Text eine Präposition und ein Artikel verschmolzen gefolgt vom Artikel, werden Präposition und Artikel getrennt geschrieben.

*Ctok*: Ich mag nicht schlafen im eine Zelt.

*TH1*: Ich mag nicht in einem Zelt schlafen.

Verschmelzungen aus Präposition und Artikel: Im mündlichen Sprachgebrauch übliche Verschmelzungen von Präposition und Artikel werden nicht korrigiert. (z.B. «durchs»)

### TOKEN LÖSCHEN

Fälschlicherweise verdoppelte Sequenzen werden rechts gelöscht.

tok	und	ich	und	ich	bin	sehr	freundlich
TH1	und	ich			bin	sehr	freundlich
-TH1			und	ich	bin	sehr	freundlich

### VERÄNDERN > HINZUFÜGEN > LÖSCHEN

Spricht keine der vorangegangenen Regeln dagegen, wird wenn möglich verändert, bevor Token hinzugefügt respektive gelöscht werden.

tok	Aber	der	Hovawart	und	der	Labrador	mag	ich	am	liebsten	.
TH1	Aber	den	Hovawart	und	den	Labrador	mag	ich	am	liebsten	.
-TH1	Aber		Hovawart	und		Labrador	mag	ich	am	liebsten	