

Analysis of (Advanced German) Learner Language by Means of the Falko Learner Corpus Family

Common Research Aims in Second Language Studies:

- > Understanding language acquisition process
- > Detecting structural learner difficulties
- > Developing improved teaching approaches

Learner Corpus Methodology – Perspectives

(Cf. Granger 2008):

1. Contrastive Interlanguage Analysis (CIA) (comparison of grammatical features in learner populations and native speakers)
→ See examples at the right below
2. Error Analysis (EA) (analysis of ungrammatical features, specific error types etc.)
→ See examples at the left below

The Falko Family:

(all data accessible at

<https://korling.german.hu-berlin.de/annis3/falko>)

Name	Text type	Tokens	Texts	Types of LI	Annotations
Essay L2 (v2.4)	Essay	144.619	248	>30	TH1, TH2, pos, lemma, diff-tags, s
Essay L1 (v2.3)	Essay	70.615	95	de	"
WHiG L2 (v2.1)	Essay	130.949	196	en	"
Kobalt L2 (v1.4)	Essay	33.368	51	chin, swe, russ	also topolog. fields
Kobalt L1 (v1.4)	Essay	12.984	20	de	"
KanDel (long.v2)	Bericht	34.612	185	en	TH1, pos, lemma, diff
Summary	Zusammenfassung	40.638	106	>10	topolog. fields, no TH
Dulko (to app.)	Essay, Übers.	--	--	ung	+error tags
Σ		467.785			

↑ Tab. 1: Overview of publically released Falko corpora

Example 1 for Systematic Error Searches: Two Constituents in the German Prefield

↑ Fig. 1: Corpus search in Falko v.2.4. Learner text lacks an article. Search expression (top left) systematically finds this type error type.

Example 2 for Sytematic Error Searches: Ungrammatical Prepositional Object Form

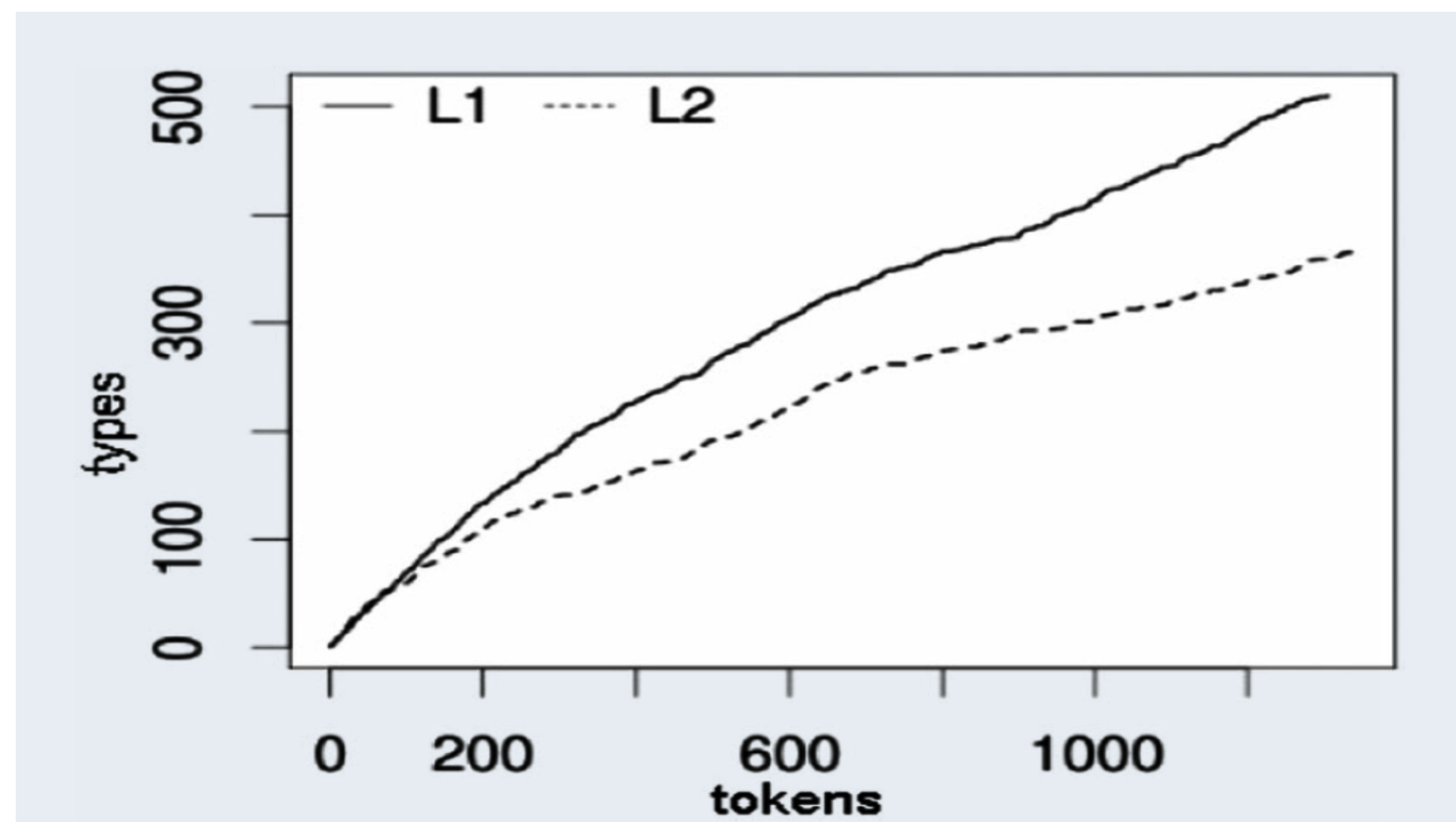
↑ Fig. 3: Corpus search in Falko v.2.4. Learner text uses erroneous prepositional form as verbal object and syntactic interpretation of corrected structure. Search expression (top left) systematically finds this type error type.

Example 1 for CIA Study: Underuse of Syntactic Categories (Hirschmann et al. 2013)

bigram	tot_norm	de	da	en	fr	pl	ru
\$.-PPER	0.042384	0.005297	0.009748	0.007963	0.006166	0.005801	0.007409
VVFIN-\$	0.042131	0.006457	0.00776	0.006343	0.006937	0.006243	0.008391
PPOSAT-NN	0.041739	0.008058	0.007247	0.007269	0.007066	0.006298	0.005802
ADV-ADV	0.041604	0.012858	0.010518	0.006111	0.006166	0.003094	0.002856
ADV-APPR	0.039742	0.009117	0.008016	0.005324	0.007837	0.004807	0.004642
PDAT-NN	0.03956	0.005409	0.004233	0.005509	0.007837	0.007735	0.008837
ADV-ART	0.037125	0.007629	0.006349	0.006898	0.005653	0.006133	0.004463

↑ Tab. 2: The frequency analysis shows the most underused word classes for Falko learners of German with different mother tongues as opposed to German native speakers: adverbs are most significantly underused.

Example 2 for CIA study: Productivity in the Use of Particle Verbs (Lüdeling et al. 2017)



↑ Fig. 2: Diverging vocabulary growth curves for particle verbs between German learners and German native speakers shows that advanced Falko learners use particle verbs less productively.

- > Granger, Sylviane (2008) Learner corpora. In: Lüdeling, Anke; Kytö, Merja (eds.) Corpus Linguistics. An International Handbook. Vol 1. Berlin; de Gruyter, S. 259-275.
- > Hirschmann, Hagen; Lüdeling, Anke; Rehbein, Ines; Reznicek, Marc; Zeldes, Amir (2013) Underuse of Syntactic Categories in Falko. A Case Study on Modification. In: 20 years of learner corpus research. Looking back, Moving ahead (LCR2013). Louvain-la-Neuve, Belgium; Presses universitaires de Louvain.
- > Lüdeling, Anke; Hirschmann, Hagen; Shadrova, Anna (2017) Linguistic Models, Acquisition Theories, and Learner Corpora: Morphological Productivity in SLA Research Exemplified by Complex Verbs in German. In: Language Learning, 1-34.