

# Contrastive analysis

---

Volker Gast

## 1 Introduction

Narrowly defined, contrastive analysis investigates the differences between pairs (or small sets) of languages against the background of similarities and with the purpose of providing input to applied disciplines such as foreign language teaching and translation studies. With its largely descriptive focus contrastive linguistics provides an interface between theory and application. It makes use of theoretical findings and models of language description but is driven by the objective of applicability. Contrastive studies mostly deal with the comparison of languages that are 'socio-culturally linked', i.e. languages whose speech communities overlap in some way, typically through (natural or instructed) bilingualism.

## 2 Contrastive analysis and foreign language teaching

Pairwise language comparison has been used in the description of foreign languages at least since the 19<sup>th</sup> century in Europe (cf. Fisiak 1981 for pertinent references). A contrastive perspective is also implicitly taken in traditional grammar writing based on the blueprint of Latin, whose linguistic system has often been superimposed on modern languages, thus implying an (asymmetrical) comparison. A contrastive methodology was explicitly formulated after the Second World War, when the importance of foreign language learning was recognized in the US, and when research on immigrant bilingualism emerged (Weinreich 1953, Haugen 1956). Charles Carpenter Fries, in his monograph on *Teaching and Learning English as a Second Language* (1945), contended that "[t]he most efficient materials are those that are based upon a scientific description of the language to be learned, carefully compared with a parallel description of the native language of the learner" (Fries 1946: 9). The programme sketched by Fries was comprehensively realized a few years later by Lado (1957) in a comparison of English and Spanish. In the preface, Lado claimed that "[...]we can predict and describe the patterns that will cause difficulty in learning, and those that will not cause difficulty, by comparing systematically the language and culture to be learned with the native language and culture of the student." This assumption has come to be known as the 'contrastive hypothesis' in its 'strong' or 'predictive' form (see e.g. Wardhaugh 1970).

The contrastive program was extensively put into practice in the 1960s, most notably with the publication of the *Contrastive Structure Series* edited by Ch. Ferguson and published by the University of Chicago Press (e.g. Moulton 1962, Kufner 1962 on English/German, Stockwell, Bowen & Martin 1965 and Stockwell & Bowen 1965 on English/Spanish). Explicit recommendations concerning the design of teaching materials and syllabi were often made, e.g. in the form of 'hierarchies of difficulty' (Stockwell et al. 1968, Stockwell & Bowen 1965). The rapid ascent of contrastive linguistics in the US culminated in the 1968 Georgetown Roundtable (Alatis 1968). While the following years witnessed a certain stagnation and even decline of contrastive analysis in the US, the discipline gathered speed in Europe, and several contrastive projects were launched, e.g. in Jyväskylä, Mannheim, Poznań,

Stuttgart and Zagreb, most of them comparing English to the native languages of the investigators (cf. Fisiak 1981).

Given that the programme of contrastive linguistics lacked a solid foundation in learning psychology, it met with empirical problems and was severely criticized before long (cf. Newmark & Reibel 1968 for a prominent critique, and James 1971 for an 'exculpation'). Originally based on a behaviourist view of language acquisition, its very foundations were called into question when behaviourism went out of fashion (cf. James 1980 for discussion). One central point of criticism was that contrastive linguistics at that stage overemphasized the role of interference as a source of errors and was too undifferentiated with respect to various parameters of language acquisition such as the question of age, the difference between mediated and natural acquisition, between the acquisition of a second or third language, etc. Even from a practical point of view, contrastive studies turned out to be of little use to language teachers, as its predictions were unreliable. Many (especially American) proponents of early contrastive analysis therefore shifted their attention to empirical studies of learners' behaviour, e.g. error analysis (Corder 1967 and subsequent work) and the study of learners' varieties, which were regarded as 'approximative systems' (e.g. Nemser 1971) or 'interlanguages' (e.g. Selinker 1972).

As the expectations of early contrastive linguistics were soon dampened, an alternative and more modest, 'weak' or 'diagnostic' form of the contrastive hypothesis was formulated (cf. Stockwell 1968, Wardhaugh 1970). It says that contrastive analyses can explain (independently) observed errors: "[T]he starting point in the contrast is provided by actual evidence from such phenomena as faulty translation, learning difficulties, residual foreign accents, and so on ..." (Wardhaugh 1970: 7–8). Accordingly, contrastive analysis and error analysis are often regarded as being complementary (cf. James 1980). Proponents of contrastive linguistics have pointed out, however, that certain discrepancies between learner varieties and the relevant target languages are not directly observable, e.g. 'covert errors' and underrepresentation or overrepresentation (cf. Rohdenburg 1974, James 1980, Kortmann 1996), and that the identification of learning difficulties thus cannot rely on error analysis alone.

While contrastive analysis in its earlier stages focused on the linguistic system, i.e. grammar and the lexicon ('microlinguistics', in James' 1980 terms), in the 1980s and 1990s matters of language use discourse structure ('macrolinguistics') came to the fore and new fields such as contrastive sociolinguistics (Hellinger and Ammon 1996), cross-cultural pragmatics (Wierzbicka 1985, 1992) and contrastive rhetoric (Connor 1996) emerged. For example, languages or perhaps speech communities differ in their degrees of directness, explicitness, etc., which may lead to intercultural misunderstandings (cf. House 1996). Similarly, conventions of text structuring may vary from one language or culture to another. For instance, there are well-known differences between the ways in which scientific texts are structured in the Continental European as opposed to the Anglo-Saxon tradition. Clyne's (1987) comparison of German and English, for example, brought to light a number of general differences which advanced learners should be aware of when writing an academic text.

Even though the early didactic aspirations of contrastive analysis were too ambitious, in its 'modest' form contrastive linguistics constitutes one building block of research into second language acquisition among others. The problem of interference, while certainly not being the only reason for learning difficulties, can hardly be denied (cf. Weinreich 1953; Gass & Selinker 1983; Odlin 1989, 1996), and systematic and general descriptions of interference phenomena are best provided on the

basis of contrastive analyses. Moreover, contrastive linguistics plays an important role in the creation of language awareness (cf. Kortmann 1996, James 2005, Mair 2005). In particular, advanced learners can profit from a direct comparison of their native language with the language to be learned, thus making their implicit knowledge of the differences explicit. Contrastive linguistics has therefore been integrated into teacher training programmes at many universities, and course materials have been designed specifically for university level teaching (cf. König & Gast 2009).

### 3 Contrastive linguistics, linguistic theory and language typology

While most of the early contrastive studies such as the grammars of the *Contrastive Structure Series* were concerned with major areas of ‘core grammar’, in particular phonology and syntax, a number of more specific studies were published in the 1970s and 1980s (e.g. König 1971, Rohdenburg 1974, Plank 1984). These studies were characterized by more specific topics and a high level of granularity as well as an increasingly theoretical orientation. Applicability in second language acquisition was a desirable, yet secondary aim, and the focus was shifted to the more abstract question of why languages differ in the way they do. This epistemological objective was closely related to the programme of linguistic typology, and contrastive linguistics came to be regarded as a “limiting case of typological comparison” (cf. König 1992, 1996, forthcoming), or a “junior partner” of linguistic typology (Kortmann 1996). While in linguistic typology many languages are compared with respect to a single variant property, contrastive studies compare only two (or very few) languages, but take a broad range of phenomena into account, ideally all areas of grammar. Even though the distinction made by Fisiak (1971, 1981), among others, between ‘theoretical’ and ‘applied’ contrastive analysis is probably not categorical, it provides a useful dichotomy for the characterization of the main objectives pursued by contrastive linguists. The typologically inspired research as sketched above is certainly located towards the theoretical pole of the continuum, but it has been shown to be not entirely irrelevant to language teaching (cf. Mair 2005 on ‘spin-offs for language teachers’). For example, the typological notion of ‘markedness’ has played an important role in research on language acquisition (cf. Kortmann 1996, James 2005), and ‘typological distance’ has been claimed to correlate with difficulty of learning (cf. Kortmann 1996, referring to Schachter 1974 on the acquisition of relative clauses).

One of the central observations made by typologically oriented contrastive linguists is that “certain differences between two languages are connected with or even a consequence of other differences” (König 1971: 13–14). The research programme emanating from this observation culminated in John Hawkins’ monograph on *A comparative typology of English and German – Unifying the contrasts*, published in 1986. Hawkins made an “attempt to consider two whole languages from a typological-universal point of view”, searching “for unifying generalisations that underlie the variation between the major portions of the whole languages” (Hawkins 1986: 3–4). In other words, he envisaged a holistic approach to pairwise language comparison, reducing the manifold contrasts between two languages to one core difference. The generalization that Hawkins arrived at is summarized as follows: “Where the grammars of English and German contrast, the surface forms (morphological and syntactic) of German are in closer correspondence with their associated meanings ...” Hawkins (1986: 121).

Given the overarching nature of Hawkins’ (1986) generalization, it comes as no surprise that counter examples were soon pointed out, and some of the more specific claims made by Hawkins were

relativized (e.g. Mair 1990, Kortmann & Meyer 1992, Rohdenburg 1992). In fact, the very idea of a holistic approach to language comparison was not seriously pursued after Hawkins (1986). Nevertheless, Hawkins' study has had a great impact on the field of contrastive linguistics up to the present day. One reason is that many of the individual generalizations made by Hawkins are still valid. Moreover, Hawkins' contrastive work in the 1980s gave rise to one of the most comprehensive and most widely accepted theories of language processing (Hawkins 1992, 1994, 2004). What this shows is that contrastive linguistics with a strongly theoretical orientation can function as a 'laboratory' for the development of new theories and paradigms by combining the fine-grained analysis of language-particular data with findings from cross-linguistic work (cf. also van der Auwera forthcoming on contrastive studies as "pilot studies in typology").

## 4 Towards a corpus-based contrastive linguistics

Given that contrastive linguistics has increasingly focused on matters of performance, thus requiring attested and, ideally, quantitative data, the use of multilingual corpora has become an integral part of this discipline in the past two decades. Two major types of multilingual corpora can be distinguished (cf. Granger 2003a, Aijmer 2008): (i) corpora consisting of original texts and their translations, and (ii) corpora containing original texts from different languages representing similar registers. Corpora of type (i) are often called 'parallel corpora', those of type (ii) 'comparable corpora' (though the terminologies vary; corpora of type (i) are also sometimes called 'translation corpora'; a different terminology is used in translation studies; cf. Granger 2003a).

Parallel corpora come in various types. In the simplest case, a parallel corpus contains a (collection of) text(s) from the original language and its translation into the target language. Such a corpus could be called a 'unidirectional translation corpus'. A 'bidirectional translation corpus' was compiled by researchers of the Yugoslav Serbo-Croatian-English contrastive project directed by R. Filipović at the University of Zagreb (cf. Filipović 1984, Johansson 2003). It contained a section of the Brown-corpus translated into (what was then called) Serbo-Croat, and a comparable amount of Serbo-Croatian text translated into English. Finally, there are also 'multidirectional translation corpora'. The most prominent example of this type is the Oslo Multilingual Corpus, which contains original and translated texts from English, Norwegian, French and German (cf. Johansson 2007). The most comprehensive parallel corpus available for European languages, the Europarl corpus, is made up of speeches given in the European Parliament and their translations into the other official languages of the European Union (cf. Koehn 2005).

Comparable corpora are well known outside the realm of contrastive linguistics, e.g. in the study of English varieties. For example, the Brown/LOB-corpora as well as the ICE-corpora contain comparable text fragments from different varieties of English (cf. Xiao 2008), thus allowing for a cross-variety comparison of specific linguistic features. The structure of the Brown family of corpora was also replicated in the (German) Limas corpus compiled at the University of Bonn, which can thus be used for English-German comparison. A number of more specific comparable corpora have become available in the recent past, e.g. the Tübingen treebanks of German, English and Japanese. For more examples of comparable corpora, the reader is referred to Aijmer (2008).

A special type of corpus is 'bilingual' or 'multilingual' in a different sense: Learner corpora contain material from only one language, but the language represents non-native and hence, in a sense,

'bilingual' output (cf. Gast forthcoming). One of the most important learner corpora, the International Corpus of Learner English, has been compiled at the University of Louvain under the direction of S. Granger (cf. Granger 2003b). Access to large amounts of learner data provides quantitative research in the domain of second language acquisition with a new dimension and moreover opens up new possibilities for the use of corpora in the classroom, e.g. as envisaged in 'data-driven learning' (see for instance Hunston 2002, Römer 2008).

While a substantial amount of multilingual corpus resources is now available, the methods of statistical analysis in this domain are still under development. In order to obtain interesting results from corpus studies, a canon of methods and procedures is needed, just as in language-particular studies (e.g. Gries 2009). Contrastive corpus studies – if they want to go beyond a simple comparison of frequencies in a pair or set of corpora – require an elaborate research design, and the problems addressed are typically multi-factorial. In the simpler cases, the methods used in language-particular analysis – e.g. cross-tabulation and significance testing – can be transferred to contrastive linguistics (e.g. Gast & Wiechmann 2011 on cleft sentences in English and German). More specific research questions, however, require a more sophisticated methodology using up-to-date statistical machinery (e.g. Wiechmann 2011). Empirically oriented contrastive linguists will thus face two challenges in the near future: developing this methodology further, and making it accessible to a broad community of researchers that is not so far familiar with quantitative methods of analysis.

## References

- Aijmer, K. (2008). Parallel and comparable corpora. In Lüdeling and Kytö, 275–291.
- Alatis, J. (ed.) (1968). *Report of the 19<sup>th</sup> Annual Round Table Meeting on Linguistics and Language Studies: Contrastive Linguistics and its Pedagogical Implications*. Monograph Series on Language and Linguistics 21. Washington: Georgetown University Press.
- Clyne, M. (1987). Cultural differences in the organization of academic texts – English and German. *Journal of Pragmatics* 11: 211–247.
- Connor, U. (1996). *Contrastive Rhetoric: Cross-Cultural Aspects of Second Language Writing*. Cambridge: Cambridge University Press.
- Corder, P. (1967). The significance of learners' errors. *International Review of Applied Linguistics* 5.4: 161–169.
- Filipović, R. (1984). What are the primary data for contrastive analysis? In Fisiak, J. (ed.), *Contrastive Linguistics*, 107–108. Berlin: Mouton de Gruyter.
- Fisiak, J. (1971). The Poznań Polish-English contrastive project. In Filipović, R. (ed.), *Zagreb Conference on English Contrastive Projects*, 87–96. Zagreb: University of Zagreb.
- Fisiak, J. (ed.) (1981). *Contrastive Analysis and the Language Teacher*. Oxford: The Pergamon Institute of English.
- Fries, C.C. (1945). *Teaching and Learning English as a Second Language*. Ann Arbor: University of Michigan Press.

- Gass, S. and L. Selinker (eds.) (1983). *Language Transfer in Language Learning*. Rowley, MA: Newbury House.
- Gast, V. (forthcoming). Contrastive analysis: Theories and methods. In Kortmann, B. and J. Kabatek (eds.), *Dictionaries of Linguistics and Communication Science: Linguistic theory and methodology*. Berlin: Mouton de Gruyter.
- Gast, V. & D. Wiechmann (2011). W(h)-clefts im Deutschen und Englischen. Eine quantitative Untersuchung auf Grundlage des Europarl-Korpus. In Gunkel, L. and G. Zifonun (eds.) *Jahrbuch des IDS 2011*. Berlin: de Gruyter Mouton.
- Granger, S. (2003a). The corpus approach: A common way forward for contrastive linguistics and translation studies? In Granger et al., 17–29.
- Granger, S. (2003b). The international corpus of learner English: A new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly* 37.3: 538–545.
- Granger, S., J. Lerot, S. Petch-Tyson (eds.) (2003). *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam/Atlanta: Rodopi.
- Gries, S. (2009). *Statistics for Linguistics with R: A Practical Introduction*. Berlin: de Gruyter Mouton.
- Haugen, Einar (1956). *Bilingualism in the Americas. A Bibliography and Research Guide*. American Dialect Society. Alabama: University of Alabama Press.
- Hawkins, J. (1986). *A Comparative Typology of English and German: Unifying the Contrasts*. London: Croom Helm.
- Hawkins, J. (1992). A performance approach to English/German contrasts. In Mair and Markus, 115–136.
- Hawkins, J. (1994). *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
- Hawkins, J. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hellinger, M. & U. Ammon (1996). *Contrastive Sociolinguistics*. Berlin: Mouton de Gruyter.
- House, J. (1996). Contrastive discourse analysis and misunderstanding: The case of German and English. In Hellinger & Ammon, 345–361.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- James, C. (1971). The exculpation of contrastive linguistics. In Nickel, G. (ed.), *Papers in Contrastive Linguistics*, 53–68. Cambridge: Cambridge University Press.
- James, C. (1980). *Contrastive Analysis*. London: Longman.
- James, C. (2005). Contrastive analysis and the language learner. In David J. Allerton, Cornelia Tschichold, and Judith Wieser (eds.), *Linguistics, Language Teaching and Language Learning*, 1–20. Basel: Schwabe.

- Johansson, S. (2003). Contrastive linguistics and corpora. In Granger et al., 31–43.
- Johansson, S. (2007). *Seeing through Multilingual Corpora: On the use of Corpora in Contrastive Studies*. Amsterdam: Rodopi.
- Koehn, Philipp (2005): *Europarl: A parallel corpus for statistical machine translation*. Machine Translation Summit X, Phuket, Thailand, 79–86.
- König, E. & V. Gast (2009). *Understanding English-German Contrasts*. 2<sup>nd</sup> ed. Berlin: Erich Schmidt Verlag.
- König, E. (1971). *Adjectival Constructions in English and German. A Contrastive Analysis*. Heidelberg: Julius Groos.
- König, E. (1992). Contrastive linguistics and language typology. In Mair & Markus, 137–154.
- König, E. (1996). Kontrastive Grammatik und Typologie. In Lang, E. and G. Zifonun (eds.), *Deutsch – typologisch*. Berlin & New York: Mouton de Gruyter.
- König, E. (forthcoming). The place of contrastive linguistics in language comparison. *Languages in Contrast*.
- Kortmann, B. (1996). Kontrastive Linguistik und Fremdsprachunterricht. In Börner, W. and K. Vogel (eds.), *Kontrast und Äquivalenz. Beiträge zu Sprachvergleich und Übersetzung*, 136–167. Tübingen: Narr.
- Kortmann, B. and P.G. Meyer (1992). Is English grammar more explicit than German grammar after all? In Mair and Markus, 155–166.
- Kufner, H.L. (1962). *The Grammatical Structures of English and German*. Chicago: University of Chicago Press.
- Lado, R. (1957). *Linguistics across Cultures*. Ann Arbor: University of Michigan Press.
- Lüdeling, A. and M. Kytö (eds.) (2008). *Corpus Linguistics. An International Handbook. Vol. I*. Berlin: de Gruyter Mouton.
- Mair, C. (1990). A contrastive analysis of object control in English and German. *Papers and Studies in Contrastive Linguistics* 25: 85-101.
- Mair, C. (2005). Recent Advances in Contrastive Linguistics and Language Typology: The Spin-Off for Language Teachers. In David J. Allerton, Cornelia Tschichold, and Judith Wieser (eds.), *Linguistics, Language Teaching and Language Learning*, 21–39 Basel: Schwabe.
- Mair, C. and M. Markus (eds.) (1992). *New Departures in Contrastive Linguistics*. Innsbruck: University of Innsbruck Press.
- Moulton, W.G. (1962). *The Sounds of English and German*. Chicago: University of Chicago Press.
- Nemser, W. (1971). Approximative systems of foreign language learners. *International Review of Applied Linguistics* 9: 115–123.

- Newmark, L. & D.A. Reibel (1968). Necessity and sufficiency in language learning. *International Review of Applied Linguistics* 6.3: 145–164.
- Odlin, T. (1989). *Language Transfer: Cross-linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
- Odlin, T. (1996). On the recognition of transfer errors. *Language Awareness* 5.3/4: 166–179.
- Plank, F. (1984). Verbs and objects in semantic agreement. Minors differences that might suggest a major one. *Journal of Semantics* 3: 305–360.
- Rohdenburg, G. (1974). *Sekundäre Subjektivierungen im Englischen und Deutschen: Vergleichende Untersuchungen zur Verb- und Adjektivsyntax*. Bielefeld: Cornelsen-Velhagen und Klasing.
- Rohdenburg, G. (1992). Bemerkungen zu infinite Konstruktionen im Englischen und Deutschen. In Mair & Markus, Vol. I, 187–207.
- Römer, U. (2008). Corpora and language teaching. In Lüdeling & Kytö, 112-130.
- Schachter, J. (1974). An error in error analysis. *Language Learning* 24: 205–214.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics* 10: 209–231.
- Stockwell, R.P. (1968). Contrastive analysis and lapsed time. In Alatis, 11–36.
- Stockwell, R.P. & J.D. Bowen (1965). *The Sounds of English and Spanish*. Chicago: University of Chicago Press.
- Stockwell, R.P., J.D. Bowen and J.W. Martin (1965). *The Grammatical Structures of English and Spanish*. Chicago: University of Chicago Press.
- van der Auwera, Johan (forthcoming). From contrastive linguistics to linguistic typology. *Languages in Contrast*.
- Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly* 4.2: 123–130.
- Weinreich, U. (1953). *Languages in Contact*. The Hague: Mouton.
- Wiechmann, Daniel. 2011. Exploring probabilistic differences between genetically related languages. *Languages in Contrast* 11.2: 193–215.
- Wierzbicka, A. (1985). Different cultures, different languages, different speech acts. *Journal of Pragmatics* 9.2/3: 145–178.
- Wierzbicka, A. (1992). *Cross-Cultural Pragmatics: The Semantics of Human Interaction*. Berlin: de Gruyter.
- Xiao, R. (2008). Well-known and influential corpora. In Lüdeling & Kytö, 383–456.