

## **An inverted loanword dictionary of German loanwords in the languages of the South Pacific**

Stefan Engelberg

Institut für Deutsche Sprache & University of Mannheim

*The paper reports on a dictionary of German loanwords in the languages of the South Pacific that is compiled at the Institut für Deutsche Sprache in Mannheim. The loanwords described in this dictionary mainly result from language contact between 1884 and 1914, when the German empire was in possession of large areas of the South Pacific where overall more than 700 indigenous languages were spoken. The dictionary is designed as an electronic XML-based resource from which an internet dictionary and a printed dictionary can be derived. Its printed version is intended as an ‘inverted loanword dictionary’, that is, a dictionary that – in contrast to the usual praxis in loanword lexicography – lemmatizes the words of a source language that have been borrowed by other languages. Each of the loanwords will be described with respect to its form and meaning and the contact situation in which it was borrowed. Among the outer texts of the dictionary are (i) a list of all sources with bibliographic and archival information, (ii) a commentary on each source, (iii) a short history of the language contact with German for each target language, and perhaps (iv) facsimiles of source texts. The dictionary is supposed to (i) help to reconstruct the history of language contact of the source language, (ii) provide evidence for the cultural contact between the populations speaking the source and the target languages, (iii) enable linguistic theories about the systematic changes of the semantic, morphosyntactic, or phonological lexical properties of the source language when its words are borrowed into genetically and typologically different languages, and (iv) establish a thoroughly described case for testing typological theories of borrowing.*

### **1. The concept of an inverted loanword dictionary**

Dictionaries of foreign words or loanword dictionaries contain words that are borrowed from a source language into a target language. With hardly any exceptions, dictionaries of this sort lemmatize the lexemes of the target language and submit them to an alphabetic macrostructural ordering, for example, the ‘Deutsches Fremdwörterbuch’ (1995ff). One of the few exceptions is Görlach’s (2001) ‘Dictionary of European Anglicisms: a Usage Dictionary of Anglicisms in Sixteen European Languages’, which approaches the borrowing relation from the source language English and traces the path of the English words into other European languages. A special case is the ‘World Loanword Database’ by Haspelmath & Tamor (2010) that contains loanwords in 41 languages and allows to approach the loanword relation from the target as well as from the source language. However, there is no dictionary that could be called a general inverted loanword dictionary, that is, a dictionary that lemmatizes all those words of a source language that have been borrowed by other languages. An inverted loanword dictionary would provide a description of the historic borrowing processes, trace the chain of borrowing processes from the source language into the target languages, document the intermediary stages, and capture the phonological, morphosyntactic and semantic changes the source words underwent.

A dictionary of this sort would serve different purposes. It would (i) help to reconstruct the history of language contact of the source language, (ii) provide evidence for the cultural contact between the populations speaking the source and the target languages, (iii) enable linguistic theories about the systematic changes of the semantic, morphosyntactic, or phonological lexical properties of the source language when its words are borrowed into genetically and typologically different languages, and (iv) establish a thoroughly described case for testing typological theories of borrowing (cf. also Haspelmath 2008).

## 2. German loanwords in the South Pacific

Some aspects of the history of German language contact have been intensively investigated, for example, the language contact between German and the Baltoslavic languages or between German and English. Several dictionaries and monographs about these languages document lexical borrowings from German, for example, into Serbo-Croatian (Striedter-Temps 1958), Estonian (Hinderling 1981), English (Pfeffer & Cannon 1994), the Teschen dialect of Polish (Menzel & Hentschel 2005), Czech (Newerkla 2004), and Polish (Vincenz & Hentschel in press). Other parts of this story still have to be written. In particular, the language contact in the former German colonies in Africa, Asia, and Oceania has found much less attention. This is all the more regrettable since the time between 1884 and 1914 was the only time when the German language came into contact with a large number of non-European languages within a short span of time. A comparison between colonial maps (e. g., *Deutscher Kolonialatlas mit Jahrbuch 1908*) and language maps from *Ethnologue* (Lewis 2009) results in a number of 1200 to 1300 languages that were spoken in areas that were at one time or another during that period part of the German colonial empire. Of the several aspects of this language contact situation, lexical borrowing processes from German into the indigenous languages have got particularly little attention. The project described in this paper, located at the Department of Lexical Studies at the Institut für Deutsche Sprache in Mannheim, is primarily concerned with the role of the German language in the former South Pacific colonies. It aims at (i) collecting lexical borrowings from German into the languages of the South Pacific, (ii) establishing which of these loans are direct loans and which are mediated by other languages (in particular by Tok Pisin, the major pidgin in what is now Papua New Guinea), (iii) describing the linguistic processes and sociohistorical circumstances of these processes, and (iv) explaining the kind and number of borrowing processes documented. A central task of the project is the compilation of an inverted dictionary of German loanwords in the languages of the South Pacific.

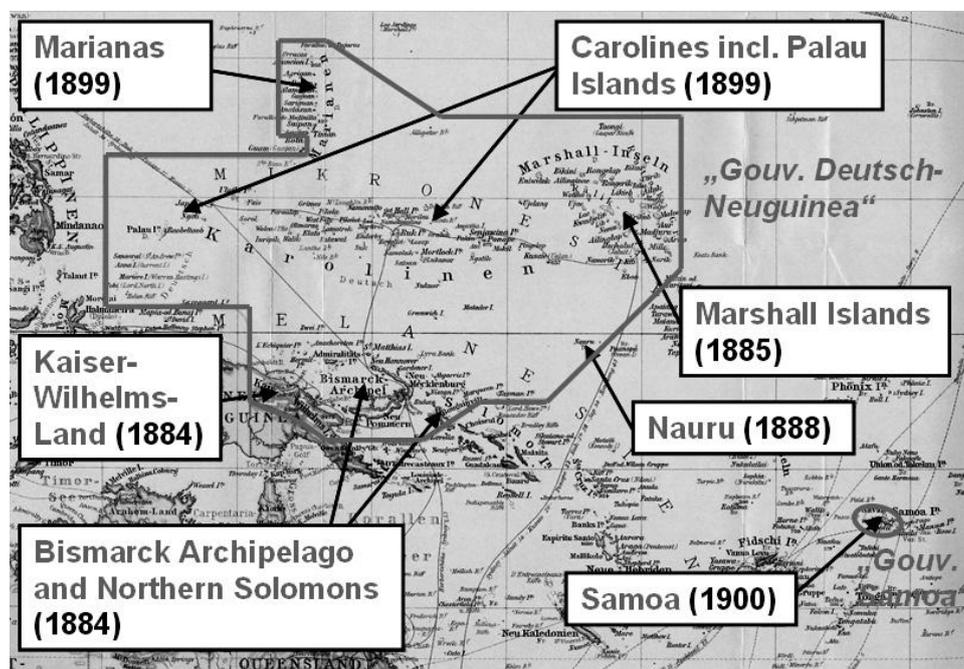


Figure 1. Dates of acquisition of German colonies in the South Pacific.

I will briefly sketch the linguistic and historic background of this project (cf. also Engelberg 2006b, 2008). Between 1884 and 1900, the German empire came into possession of large

areas of the South Pacific, either by occupation, purchase, or negotiation with other colonial powers. All of these colonies were occupied by other nations in the first year of WW I (cf. Figure 1).

More than 700 languages were spoken in the South Pacific colonies, 95% of them in the German part of New Guinea ('Kaiser-Wilhelms-Land'), the Bismarck Archipelago, and the Northern Solomons. These languages belong to different families, partly to the Oceanic branch of Austronesian – including Micronesian, Polynesian, and Western Oceanic languages – and partly to different Papuan language families: Amto-Musan, Arafundi, Arai-Kwomtari, Border, East-New Britain, Left May, Mongol-Langam, North Bougainville Pauwasi, Piawi, Ramu-Lower Sepik, Senagi, Sepik, Sko, South Bougainville, Torricelli, Trans-Newguinea, Yele-West New Britain, and Yuat (classification after Ethnologue, Lewis 2009). There were probably not more than 2000 Germans in the South Pacific colonies at one time. Thus, actual contact between Germans and the indigenous population was more or less limited to the central stations of the German administration, the German plantations, the few other German enterprises (e. g., phosphate mines), and the stations run by the German missions. However, German was put into the curricula of about 750 schools, five of which were run by the government and the others by the missions (cf., e. g. Schlunk 1914). Engelberg (2006a) shows that the quality of the German lessons differed considerably depending on whether German was also the language of instruction, whether the school was run by a German or non-German mission, and whether the teacher was a native speaker of German. Anyway, many indigenous pupils had at least some exposure to the German language. The language contact between German and the indigenous languages resulted in (i) the origin of German-based pidgins and creoles like Ali Pidgin German (Mühlhäusler 1979a) and Unserdeutsch (Volker 1991), (ii) a heavy lexical influence of German on Tok Pisin (Mühlhäusler 1979b), and (iii) borrowing processes from German into the indigenous languages, e. g.:

- *Amtmann* 'senior civil servant' > Samoan *'ametimani* (Heider 1913)
- *Arbeit* 'work' > Buin *arapaita* (Laycock 1971)
- *deutsch* 'German' > Mokilese *dois* (Harrison 1977)
- *Gummi* 'rubber' > Trukese *kkumi* (Goodenough 1980)
- *malen* 'to paint' > Marshallese *malen* (Abo 1976)
- *Papier* 'paper' > Jabêm *papia* (Streicher 1937)
- *Schrank* 'cupboard' > Palauan *sérangk* (McManus 1977)
- *Spaten* 'spade' > Takia *spaten* (Ross 2009)
- *Tafel* 'blackboard' > Gedaged *tafe* (Fischer 2000)

A small number of German loans can also be found in other areas of the South Pacific, either transported there via Tok Pisin (e. g., German *Besen* 'broom' into Dehu *beisin*) or they are a result of pre-colonial contact with German speaking traders or explorers (e. g., German *hanseatisch* 'hanseatic' into Hawaiian *haneseatika* or, occasionally, German *Eisenbahn* 'train, railway' into Maori *aihanapana*).

### 3. The structure of the dictionary

*Dictionary medium:* The dictionary is designed as an electronic resource from which an internet dictionary and a printed dictionary can be derived. Technically, the dictionary will be realized in XML format. A dictionary specific document-type definition that expresses the structural properties of the XML-based dictionary is being designed. The data will be held in an Oracle database and be published in a dictionary portal connected to and based on the

principles of the OWID portal at the Institut für Deutsche Sprache (cf. Müller-Spitzer 2008, Engelberg, Klosa & Müller-Spitzer 2009).

*Dictionary basis:* The dictionary basis consists of (i) primary sources from the turn of the century, written in the target languages (religious publications by the missions, private letters, administrative or legal texts, and textbooks produced for the local schools), (ii) secondary sources (in particular, dictionaries and word lists of the languages of the South Pacific), and (iii) tertiary sources (travelogues, mission reports, missionary correspondence, local German newspapers, archival documents from the German administration, etc.). So far, more than 2000 source texts have been assembled.

*Macrostructure:* The lexical basis of the dictionary is a list of pairs of German source words and target words in the languages of the South Pacific. So far, more than 600 loanwords in almost 30 languages have been collected (cf. Figure 2). Some of the loanwords are still in use; others have meanwhile been replaced by loans from other languages, or they have hardly caught on in the first place. Thus, the current status of the loanword is not decisive for its being included in the dictionary.

German etymon	loanword	target language	sources
...	...	...	...
<i>Mahlzeit</i> ‘meal’	<i>malsait</i>	Tok Pisin	Mühlhäusler (1979)
<i>Mai</i> [?] ‘May’	<i>mai</i>	Bongu	Hanke (1909)
<i>Mai</i> [?] ‘May’	<i>Mai</i>	Jabêm	Streicher (1937)
<i>Mais</i> [?] ‘corn’	<i>mais</i>	Palauan	HLM-008
<i>Majestät</i> ‘majesty’	<i>maiesitete</i>	Samoaan	Lynch (2004), HLP-058
<i>Malaria</i> [?] ‘malaria’	<i>malaria</i>	Tok Pisin	Steinbauer (1969)
<i>malen</i> ‘paint’	<i>malen</i>	Marshallese	Abo (1976), Lynch (2004)
<i>malen</i> ‘paint’	<i>mahlen, mahlenih, maleniedi, sounmalen, mahmahlen</i>	Ponapean	Rehg (1979), Burdick (1970)
<i>malen</i> ‘paint’	<i>malen</i>	Tok Pisin	Mühlhäusler (1979)
<i>Manna</i> ‘manna’	<i>(manna)</i>	Nauruan	Gründl (1906)
<i>Mark</i> ‘Mark’	<i>maak, mak</i>	Marshallese	Abo (1976), Lynch (2004), HLM-175
<i>Mark</i> ‘Mark’	<i>mak</i>	Palauan	McManus (1977), ML-041
<i>Mark</i> ‘Mark’	<i>Mark</i>	Ponapean	Hahl (1904)
<i>Mark</i> ‘Mark’	<i>maak, makey</i>	Puluwat	Elbert (1972)
<i>Mark</i> ‘Mark’	<i>maka, mark, mareka, Maka, mareke</i>	Samoaan	[Solf 1905], Lynch (2004), Mader (o.J.), NL-008, HLP-058
<i>Mark</i> ‘Mark’	<i>mak</i>	Tok Pisin	Mühlhäusler (1979), Steinbauer (1969)
<i>Mark</i> ‘Mark’	<i>maak</i>	Trukese	Goodenough (1980)
<i>Mark</i> ‘Mark’	<i>mak</i>	Ulithian	Walsh (1979)
<i>Mark</i> ‘Mark’	<i>maak</i>	Woleaian	Sohn (1976)
<i>Marsch</i> ‘march’	<i>maas</i>	Carolinian	Jackson (1991)
<i>März</i> [?] ‘March’	<i>Melese</i>	Jabêm	Streicher (1937)
<i>Maschine</i> ‘machine’	<i>mesil, masil</i>	Palauan	McManus (1977), HLM-008
<i>Maschinengewehr</i> ‘machine gun’	<i>mesilkebiér</i>	Palauan	McManus (1977), Josephs (1984)
...	...	...	...

Figure 2. Excerpt from the current list of borrowings.

[The ‘?’ expresses doubt about the etymological information found in the sources. The sources comprise of published texts and documents from the archives. (They are not listed in the reference section of this article.)]

The outer access structure of the printed dictionary will be based on an alphabetic list of the German source words. The internet version will allow access to the dictionary entries by

searching for the source words, the target words, the languages borrowed into, and other criteria.

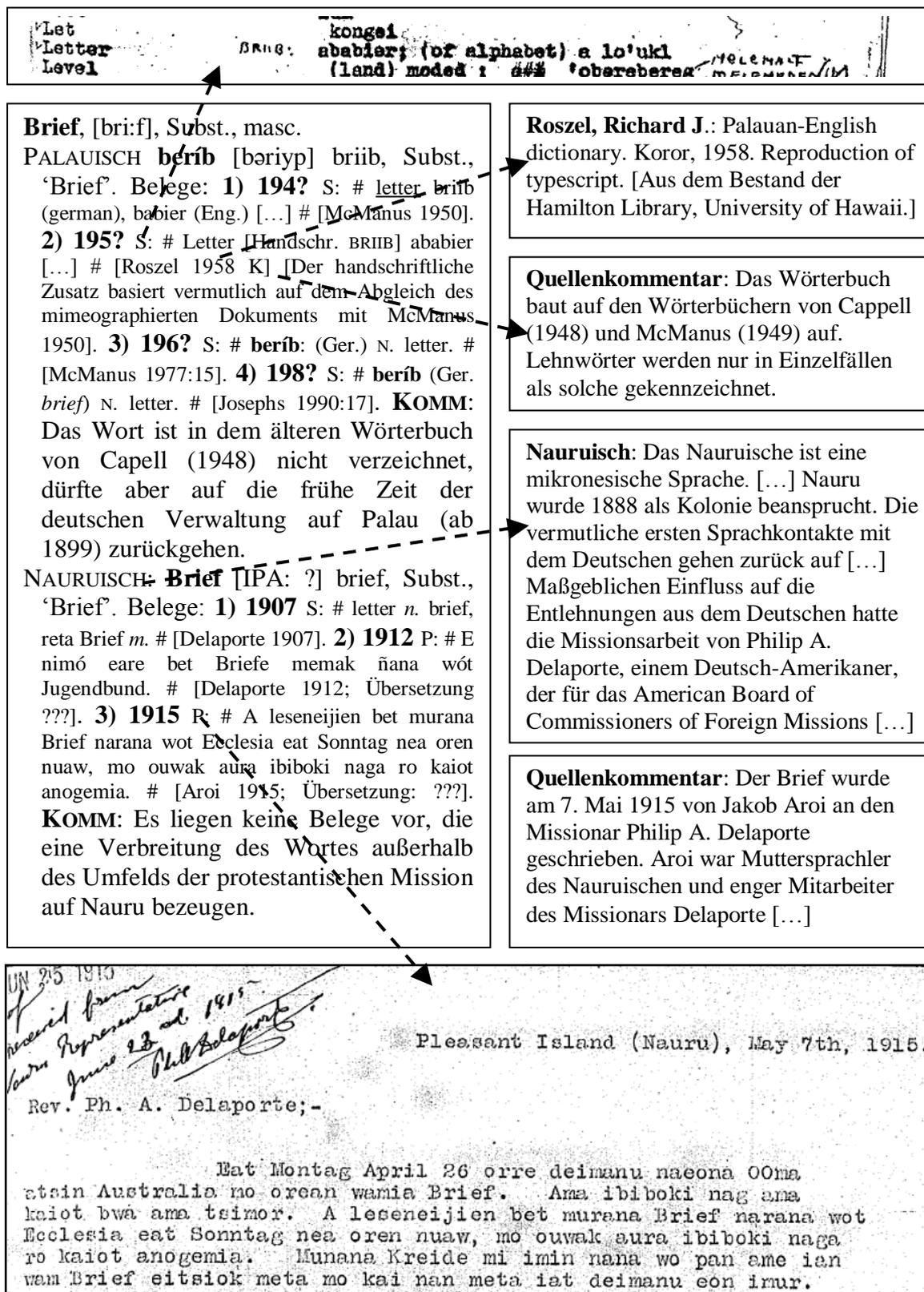


Figure 3. Example for a dictionary article and its links to the outer texts of the dictionary.

*Microstructure:* Figure 3 illustrates how a printed version of a dictionary article could look. The internet version will allow different user-adaptive views on the data according to the principles of the dictionary portal OWID (cf. Müller-Spitzer 2008, Engelberg & Müller-Spitzer in print). The information positions include (1) the German source word, (2) a form comment (pronunciation, morphosyntactic properties), and for each loanword (3) a subsection in which the loanword is described. Each subsection comprises of (3.1) the target language, (3.2) the loanword, (3.3) a form comment (pronunciation, orthographic variants, morphosyntactic properties), (3.4) a semantic comment containing a meaning description of the loanword and a list of pieces of evidence, and (3.5) a commentary. The evidence section contains evidence from primary (P), secondary (S), and tertiary (T) sources. Each evidence position is supplied with a date, the actual text from the source, the reference to the list of the sources, and an optional commentary. The date is not necessarily the date of the publication but indicates when, according to the source, the word was encountered. With dictionaries as secondary sources, the actual data collection often precedes the publication of the dictionary by many years. The source text can be a primary text; in which case, a translation is given (which is still missing in the example article). In other cases, excerpts from dictionaries, grammars, or other sources are integrated.

*Outer texts:* Except for the usual outer texts of the dictionary such as the scientific introduction, the instructions for use, and the list of abbreviations, four particular kinds of texts will supplement the dictionary: (i) a list of all sources with bibliographic and archival information, (ii) a commentary on each source (K), (iii) for each target language a short history of the language contact with German, and perhaps (iv) facsimiles of source texts. (The last point has not finally been decided.) The commentary will help to judge the value of the source as an indicator for the borrowing processes. For example, with some primary sources, it is not clear whether they were written by German missionaries or by or at least with the help of native speakers. Dictionaries are more or less reliable with respect to the attribution of loanwords to German vs. English origin, and comments on German loanwords in travelogues have to be carefully evaluated. The short contact history given for each language will reveal the basic facts of the sociohistorical context of the borrowing processes. In the internet version of the dictionary, links are provided from the article to the outer texts (cf. Figure 2).

#### 4. Challenges

With respect to the lexicographic practice, the dictionary differs considerably from loanword dictionaries that document language contact between two languages with a long history of written records. Firstly, reliable data are hard to come by. Written documents as primary sources from the time around the turn of the century are scarce and only available for very few languages. Secondary sources such as dictionaries compiled during the 20<sup>th</sup> century are available for some languages, in particular those of Micronesia. The languages of New Guinea and other parts of Melanesia are considerably less well documented. Secondly, the lexicographers are competent in the source language but not in the target languages; wherever metalinguistic knowledge cannot be sufficiently gained from the linguistic literature, experts of the target languages will be required. However, with more than 700 languages spoken in the Government of German New Guinea, this is a very restricted option. Thirdly, finding information about loanwords and in particular about the circumstances of the borrowing processes and the historic language situation requires a great deal of archival work which is not facilitated by the fact that the relevant archives are located far apart from each other (Canberra, Berlin, Honolulu, Wellington, Boston, Samoa, etc.). Thus, we have to allocate an

unusually large amount of time to the acquisition and interpretation of sources compared to the time needed for the actual compilation of the dictionary.

## 5. Outlook

At the beginning of this paper, the concept of an inverted loanword dictionary of the German language was presented. The dictionary described in this article is of course only a small contribution to such a project, and it is difficult to envisage how a single group of lexicographers would ever accomplish such a task at all. On the other hand, quite a number of loanword dictionaries have been and are written from the perspective of the languages that borrowed from German. This suggests that the task should be tackled in a collaborative manner. In cooperation with other institutions, in particular the Slavic Department at the University of Oldenburg, a concept for an internet loanword dictionary portal is being developed at the Institut für Deutsche Sprache. In accord with the general concept of dictionary nets (cf. Engelberg & Müller-Spitzer in print), dictionaries documenting loans from German will be collected within an internet-based dictionary portal. The dictionaries will be available as stand-alone products, but they will also be integrated in a net-like fashion with a common metalemmalist, cross-dictionary references, and search options.

The first two dictionaries, documenting loans from Polish (Menzel & Hentschel 2005, Vincenz & Hentschel in press) are currently converted into adequate XML structures. Many conceptual problems pertaining, for example, to an ontological basis for onomasiological searches or to the adequate structure of a metalemmalist comprising historical and dialectal variants of German still have to be solved.

## References

- Abo, T.; Bender, B. W.; Capelle, A.; DeBrum, T. (1976). *Marshallese-English Dictionary*. Honolulu: University Press of Hawaii.
- Deutscher Kolonialatlas mit Jahrbuch* (1908). Ed. by Deutsche Kolonialgesellschaft. Berlin: Reimer.
- Deutsches Fremdwörterbuch* (1995ff). 12 vol., 2. ed., started by Hans Schulz & continued by Otto Basler. Berlin, New York: de Gruyter. [6 vols. published so far; 1. ed., Straßburg 1913-1983.]
- Engelberg, S. (2006a). 'The Influence of German on the Lexicon of Palauan and Kosraean' [on line]. In Allen, K. (ed.). *Selected Papers from the 2005 Conference of the Australian Linguistic Society*. <http://www.als.asn.au/proceedings/als2005.html> [Access date: 23 Jan. 2010].
- Engelberg, S. (2006b). 'Kaisa, Kumi, Karmoból. Deutsche Lehnwörter in den Sprachen des Südpazifiks'. In *Sprachreport* 22 (4). 2-9.
- Engelberg, S. (2008). 'The German Language in the South Seas: Language Contact and the Influence of Language Politics and Language Attitudes'. In Schulze, M. et al. (eds.). *German Diasporic Experiences: Identity, Migration, and Loss*. Waterloo, ON: Wilfrid Laurier University Press. 317-329.
- Engelberg, S.; Klosa, A.; Müller-Spitzer, C. (2009). 'Challenges to Internet lexicography: The Internet dictionary portal at the Institute for German Language'. In *Kernerman Dictionary News* (July). 10-19.
- Engelberg, S.; Müller-Spitzer, C. (in print). 'Dictionary Portals'. In Gouws, R. H. et al. (eds.). *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Special Focus on Computational Lexicography*. Berlin, New York: de Gruyter.
- Fischer, H. (2000). 'Hundert Jahre Wandel bei den Wampar in Papua-Neuguinea'. In *Geographische Rundschau* 52 (4). 43-38.
- Goodenough, W. H.; Sugita, H. (1980). *Trukese-English Dictionary. Pwpwuken Tettenin Fóós: Chuuk-Ingenes*. Philadelphia: American Philosophical Society.
- Görlach, M. (ed.) (2001). *A Dictionary of European Anglicisms: a Usage Dictionary of Anglicisms in Sixteen European Languages*. Oxford et al.: Oxford University Press.
- Harrison, S. P.; Albert, S. (1977). *Mokilese-English Dictionary*. Honolulu: The University Press of Hawaii.
- Haspelmath, M. (2008). 'Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability'. In Stolz, T. et al. (eds.). *Aspects of Language Contact*. Berlin: Mouton de Gruyter. 43-62.
- Haspelmath, M.; Tadmor, U. (2010). *World Loanword Database* [on line]. <http://wold.livingsources.org/> [Access date: 6 Feb. 2010].
- Heider, [P.] E. (1913). *Kurzes Vokabularium deutsch-samoanisch und samoanisch-deutsch*. Malua (Samoa): L.M.S. [London Missionary Society] Press.
- Hinderling, R. (1981). *Die deutsch-estnischen Lehnwortbeziehungen im Rahmen einer europäischen Lehnwortgeographie*. Wiesbaden: Harrassowitz.
- Laycock, D. C. (1971). 'English and Other Germanic Languages'. In Sebeok, T. A. (ed.). *Current Trends in Linguistics. Vol. 8: Linguistics in Oceania. Part two: Selected Topics*. The Hague, Paris: Mouton. 877-902.
- Lewis, M. P. (ed.) (2009). *Ethnologue: Languages of the World, Sixteenth edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com/> [Access date: 23 Jan. 2010].
- McManus, E. G. (1977). *Palauan-English Dictionary*. Edited and expanded by Lewis S. Josephs with the assistance of Masa-aki Emesiochel. Honolulu: University Press of Hawaii.
- Menzel, T.; Hentschel, G. (2005). *Wörterbuch der deutschen Lehnwörter im Teschener Dialekt des Polnischen*. Oldenburg.
- Mühlhäusler, P. (1979a). 'Bemerkungen zur Geschichte und zum linguistischen Stellenwert des 'Pidgindeutsch''. In Auburger, L.; Kloss, H. (eds.). *Deutsche Sprachkontakte in Übersee. Nebst einem Beitrag zur Theorie der Sprachkontaktforschung*. Tübingen: Narr. 59-87.
- Mühlhäusler, P. (1979b). *Growth and Structure of the Lexicon of New Guinea Pidgin*. Canberra: Australian National University, Department of Linguistics, Research School of Pacific Studies.

### Section 3. Reports on Lexicographical and Lexicological Projects

- Müller-Spitzer, C. (2008). 'The Lexicographic Portal of the IDS: Connecting Heterogeneous Lexicographic Resources by a Consistent Concept of Data Modelling'. In Bernal, E.; DeCesaris, J. (eds.). *Proceedings of Euralex XIII*. Barcelona. 457-461.
- Newerkla, S. (2004). *Sprachkontakte Deutsch - Tschechisch - Slowakisch. Wörterbuch der deutschen Lehnwörter im Tschechischen und Slowakischen: historische Entwicklung, Beleglage, bisherige und neue Deutungen*. Frankfurt/M. et al.: Lang.
- Pfeffer, J. A.; Cannon, G. (1994). *German Loanwords in English. An Historical Dictionary*. Cambridge, New York, Melbourne: Cambridge University Press.
- Ross, M. (2009). 'Loanwords in Takia, an Oceanic Language of Papua New Guinea'. In Haspelmath, M.; Tadmor, U. (eds.). *Loanwords in the World's Languages. A Comparative Handbook*. Berlin: de Gruyter. 747-770.
- Schlunk, M. (1914). *Die Schulen für Eingeborene in den deutschen Schutzgebieten am 1. Juni 1911. Auf Grund einer statistischen Erhebung der Zentralstelle des Hamburgischen Kolonialinstituts*. Hamburg: Friederichsen & Co.
- Streicher, J. F. (1937). *Wörterbuch Deutsch-Jabêm*. Neuendettelsau: Ms. [type-written].
- Striedter-Temps, H. (1958). *Deutsche Lehnwörter im Serbokroatischen*. Wiesbaden: Harrassowitz.
- Vincenz, A.; Hentschel, G. (in press). *Wörterbuch der deutschen Lehnwörter in der polnischen Schrift- und Standardsprache*. Begonnen, konzipiert und grundlegend redigiert von Andrzej Vincenz. Zu Ende geführt von Gerd Hentschel.
- Volker, C. (1991). 'The Birth and Decline of Rabaul Creole German'. In *Language and Linguistics in Melanesia* 22. 143-156.