1. The Special Interest of Suppletion

The phenomenon of suppletion, as found in English *go~went* where different inflectional forms of the same lexical item are not related phonologically, has a special place in morphology. Part of its importance is that it sets one of the outer bounds for the notion ‘possible word’ in a human language. It provokes questions about how such forms are to be treated in our theories, and how they are stored (Carstairs-McCarthy 1994). There has been considerable work on suppletion, particularly from Osthoff (1899) onwards. Current interest in the topic is shown by the recent appearance of two dissertations (Veselinova 2003 and Veselinović 2003). While the body of research is extensive, the range of languages investigated is rather restricted in many publications. In order to stimulate further progress, we have constructed and made available a database (Brown, Chumakina, Corbett and Hippisley 2004). We hope this will help to put future research on a broader empirical base. An annotated bibliography is now available (Chumakina 2004); it contains over seventy entries on works written in five different languages (English, French, German, Italian and Russian) and this will give the reader a view of the literature.

2. The Canonical Approach in Typology

At the Barcelona meeting (MMM3), the first author outlined a ‘canonical’ approach to typology. In a canonical approach, we take definitions to their logical end point and build theoretical spaces of possibilities. Only then do we ask how this space is populated. The canonical instances, which are the best examples, those most closely matching the canon, may well not be the most frequent. Rather they may be rare, or even non-existent. They serve to fix a point from which occurring phenomena can be calibrated, and it is then significant and interesting to investigate frequency distributions. This approach was worked out with regard to agreement (Corbett 2003). It is an interesting issue how such an approach can be viable for a phenomenon like suppletion, which may be thought of as an ‘extreme’ phenomenon. It is clear that the

---

1 The support of the AHRB under grant B/RG/AN4375/APN10619 and of the ESRC under grant R00027135 is gratefully acknowledged. The first author wishes to thank the organizers of the Fourth Mediterranean Morphology Meeting (MMM4), Catania, Italy, 21-23 September 2003, for the invitation to present this research, and to the participants for useful discussion. Constraints of space mean that this printed version covers only one part of the Catania presentation, namely the most collaborative part, hence it is co-authored by those involved in the Surrey Suppletion Database. At the editors’ invitation the bibliography is also included here (without the abstracts). A version of this paper was presented shortly after MMM4 at the Workshop on "Database-driven linguistic typology" at the Language Typology Research Centre Annual Meeting, Estoril.

object of such a typology will be lexemes, rather than constructions or languages. A helpful start for such an approach is offered by a part of Mel’čuk’s definition of suppletion:

For the signs X and Y to be suppletive their semantic correlation should be maximally regular, while their formal correlation is maximally irregular.

Mel’čuk (1994: 358)

Beginning from this suggestion, we can establish dimensions along which the phenomenon may vary. We can establish the canonical instances, namely those which are maximally transparent in semantic terms and maximally opaque in formal terms (cf. Mel’čuk 1994: 342). As part of this we can recognise, for example, that some restrict suppletion to inflectional morphology, while others including Mel’čuk allow for suppletion in derivational morphology. Semantic correlations are typically more regular (more transparent) in inflectional than in derivational morphology, hence the clearer (and for some linguists the only) instances of suppletion will be found in inflectional morphology. Only instances of inflectional morphology are included in the database. To date fifteen criteria for canonical suppletion have been proposed in what is ongoing research.

3. **The Surrey Database of Suppletion**

The database was designed and implemented both to inform our research and to make available to other linguists the data which we had collected and analysed. It allows a range of queries, and can be searched on-line over the web (at www.smg.surrey.ac.uk).

3.1 **Structure of the Database**

The design, due primarily to the second author, is indicated in Figure 1. Each table is motivated by a possible query.
Figure 1: Design of the Database

The design of the database in this figure allows for detailed description of the environments in which suppletion occurs and for the non-redundant storage of the information. On the right side of the figure there are the tables for feature sets (such as Case, Number, Person etc). Any feature in a feature set table occurs once in that table, but many times in the Combination table to the left of the feature sets. Feature combinations are then paired with stems (in the StemCombination table). The stem in a lexeme-stem pairing (in LexemeStem table) may occur with more than one feature combination. The relationship between the stem field in the LexemeStem table and the stem field in the StemCombination table is therefore one-to-many. The LanguageLexemeSuppletion table brings all the information together, combining the languages from the Language table with the lexemes from the LexemeStem table. The database has been implemented using Microsoft Access.

3.2 The Data

Languages were selected to ensure genetic and areal diversity. In addition, languages had to have the potential for inflectional suppletion (hence those with no inflectional morphology were not included). The data were derived from published grammars and dictionaries, and in many cases were checked with specialists. We are very grateful to Willem Adelaar, Nicholas Evans, George Hewitt, Paulette Levy, Marianne Mithun and Larry Trask for their help. The data on two languages, Komi and Xakass, were obtained
on field trips. The database records all instances of suppletion that were found in the following languages:

<table>
<thead>
<tr>
<th>Language</th>
<th>Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>!Xôô</td>
<td>Khoisan</td>
</tr>
<tr>
<td>Arapesh</td>
<td>Toricelli</td>
</tr>
<tr>
<td>Archi</td>
<td>Nakh-Daghestanian</td>
</tr>
<tr>
<td>Basque</td>
<td>Basque</td>
</tr>
<tr>
<td>Chichewa</td>
<td>Niger-Congo</td>
</tr>
<tr>
<td>Georgian</td>
<td>Kartvelian</td>
</tr>
<tr>
<td>Guarani</td>
<td>Tupi</td>
</tr>
<tr>
<td>Hebrew</td>
<td>Semitic</td>
</tr>
<tr>
<td>Hua</td>
<td>Trans-New Guinea</td>
</tr>
<tr>
<td>Hungarian</td>
<td>Ugric</td>
</tr>
<tr>
<td>Itelmen</td>
<td>Chukotko-Kamchatkan</td>
</tr>
<tr>
<td>Jacaltec</td>
<td>Mayan</td>
</tr>
<tr>
<td>Japanese</td>
<td>Japanese</td>
</tr>
<tr>
<td>Kannada</td>
<td>Dravidian</td>
</tr>
<tr>
<td>Kayardild</td>
<td>Tangkic</td>
</tr>
<tr>
<td>Ket</td>
<td>Yenisei-Ostyak</td>
</tr>
<tr>
<td>Koasati</td>
<td>Muskogean</td>
</tr>
<tr>
<td>Komi</td>
<td>Finno-Permic</td>
</tr>
<tr>
<td>Limbu</td>
<td>Sino-Tibetan</td>
</tr>
<tr>
<td>Mayali (Bininj Gun-wok)</td>
<td>Gunwinyguan</td>
</tr>
<tr>
<td>Tetercingo Nahuatl</td>
<td>Uto-Aztecan</td>
</tr>
<tr>
<td>Navajo</td>
<td>Athabaskan</td>
</tr>
<tr>
<td>Nishnaabemwin</td>
<td>Algonquian</td>
</tr>
<tr>
<td>Palauan</td>
<td>Austronesian</td>
</tr>
<tr>
<td>Qafar</td>
<td>Cushitic</td>
</tr>
<tr>
<td>Russian</td>
<td>Indo-European</td>
</tr>
<tr>
<td>Tariana</td>
<td>Arawak</td>
</tr>
<tr>
<td>Tarma Quechua</td>
<td>Quechuan</td>
</tr>
<tr>
<td>Totonac</td>
<td>Totonacan</td>
</tr>
<tr>
<td>Turkana</td>
<td>Nilo-Saharan</td>
</tr>
<tr>
<td>Xakass</td>
<td>Turkic</td>
</tr>
<tr>
<td>Yimas</td>
<td>Sepik-Ramu</td>
</tr>
<tr>
<td>Yukaghir</td>
<td>Yukaghir</td>
</tr>
<tr>
<td>Yup'ik</td>
<td>Eskimo-Aleut</td>
</tr>
</tbody>
</table>

For each example we present the phonologically distinct stems that belong to the same paradigm, and define the categories according to which the suppletion can be delineated. The database contains pointers to examples, illustrating each instance of suppletion in a particular language. In addition, there is a link to a report for each language, giving sources and enabling the user to see how decisions were made. We describe briefly for each language the morphonological processes relevant for defining suppletion, and the inflectional system (major word classes and the categories they inflect for). We list the instances of suppletion and give examples of regular inflected items for contrast. In the cases where our analysis of the language material differs from
that of the source, we present both views and give our reasons for deciding whether or not to include this particular example in the database.\textsuperscript{2}

Users can query the database online. Besides obvious searches, such as by language, it is also possible to do cross-linguistic searches in terms of semantic and morpho-syntactic categories. The web interface provides the user with pulldown menus for each of the relevant categories. There are three readme files with the database to aid initial searching.

3.3 \textbf{Some Initial Results}

A first observation is that suppletion is relatively common cross-linguistically. Out of the 34 language surveyed, in only four could we find no instances of suppletion (recall that in order to be included a language had to have inflection and hence the theoretical possibility of inflectional suppletion). The four languages are: Navajo, Tarma Quechua, Yukaghir and Yup’ik.

The database contains 178 lexical items and 417 stems. Among the morphological features involved in suppletion it is interesting to note person (in verbs) in Totonac; possession in Jacaltec and Nishnaabemwin; politeness in Japanese and Tetelcingo Nahuatl; and negation in Russian, Limbu and Hua.

It is true that the lexical items involved are usually frequent items like ‘go’ and ‘child’. But that is not invariably the case. The Nakh-Daghestanian language Archi has the following remarkable suppletive item: bič’ni (SG) / boždo (PL) ‘corner of a sack’ (Kibrik 1977: 46). Some results are presented in Hippisley, Chumakina, Corbett & Brown (2004); we intend to continue exploiting the database, in parallel with other researchers.

4. \textbf{A Bibliography of Suppletion}

A second resource relating to suppletion is the annotated bibliography whose entries we give here. The related abstracts are accessible at \textit{http://www.surrey.ac.uk/LIS/SMG/}


\textsuperscript{2}The issue of reproducibility is discussed in Corbett (2004).


Welna, J. (2001), *Suppletion for suppletion, or the replacement of êde by went in English*, “Studia Anglica Posnaniensia” 36, 95–110.


5. **Conclusion**

Suppletion is indeed a challenge for morphologists and typologists. There are some remarkable instances, which push back the boundary of what is a ‘possible word’. By constructing the database, compiling the bibliography, and making both generally available, we hope to contribute to a better understanding of this extreme phenomenon.

**References**


