

The Prosody of Short Referring Expressions in German Radio News

Arndt Riester, Kerstin Eckart and Katrin Schweitzer (IMS, University of Stuttgart)

We report on some query results from our newly established database of German radio news broadcasts (DIRNDL – Discourse Information Radio News Database for Linguistic analysis). The speech data have been annotated with intonation labels according to GToBI(S) (Mayer, 1995). Written transcripts of the news have been parsed with a state-of-the-art German LFG grammar, and labeled for *information status* following Riester et al. (2010). Both types of annotated data are linked together in a relational database management system (Eckart et al., 2010), which allows for flexible bi-directional queries. This gives us a powerful resource for studying the correspondence between prosody, morphosyntax and discourse information. We focus on a section comprising one day of hourly broadcast news (2:25 hrs, or 1193 sentences). Information status is assigned to pronouns, proper names, and determiner phrases, which may span several words. Long phrases tend to carry several pitch accents. In order to keep the data simple, we start with one-word (pronoun, proper name, bare noun) and two-word expressions (determiner-noun, preposition-noun, title-name sequences, compounds). We distinguish between coreferential terms (GIVEN, (1)), non-coreferential but context-dependent definites (BRIDGING, (2)), non-anaphoric definites (e.g. discourse-new proper names or globally unique definites; UNUSED, (3)) and discourse-new indefinites (INDEF-NEW).

(1) Der CDU-Politiker Wissmann ist zum neuen Präsidenten des VDA gewählt worden. Sein Bundestags-Mandat legt WISSmann nieder. (s1964)

CDU politician Wissmann has been elected chairman of the VDA. Wissmann will step down from his parliament mandate. (GIVEN, rising accent)

(2) Allein in der Hauptstadt Kairo sind tausende Polizisten im Einsatz, um die Volksabstimmung abzusichern. Dort war die Beteiligung am Vormittag gering. (s1736)

In the capital Cairo alone, thousands of policemen are safeguarding the referendum. There, in the morning the turnout was low. (BRIDGING, no accent)

(3) Der Fraktionschef der Linkspartei im Bundestag, Gysi, hat eine Koalition mit der SPD vorerst ausgeschlossen. (s1584)

The leader of the parliamentary group of the Left Party in the German parliament, Gysi, for the time being has ruled out a coalition with the Social Democrats. (UNUSED, no accent)

Moreover, we group the GToBI(S) accents into *falling* and *rising* ones (always: final accent on the expression). *Fall cont.* and *rise cont.* indicate a monotonal pitch accent without a trail tone. In combination with a class *no accent*, this results in five different prosodic categories. Their frequencies are shown in Tables 1-2. As expected, there are less accents on GIVEN expressions than on non-given ones; particularly, reflexive (98%) and other (80%) pronouns are for the most part unaccented. Only 12-13% the non-pronominal GIVEN items are unaccented. A GIVEN name carrying a rising (L*H) accent is shown in (1).

At the other end of the scale, we expect discourse-new items to carry an accent. This is largely confirmed. Interestingly, though, there is a certain number (7%) of BRIDGING definites (always determiner-noun) and a tiny set (2-3%) of UNUSED definites (but no discourse-new indefinites) which are unaccented, see (2-3). There is no obvious correlation between information status and accent type (rise/fall) in our data. However, for each category, rises on average occurred earlier than falling accents.

Information Status	No Accent	Rise (L*H)	Rise cont. (L*)	Fall (H*L)	Fall cont. (H*)	100%
GIVEN-REFLEXIVE	97 (98%)	--	--	1 (1%)	1 (1%)	99
GIVEN-PRONOUN	173 (80%)	29 (13%)	1 (0%)	10 (5%)	3 (1%)	196
GIVEN-NONPRONOUN	13 (13%)	41 (41%)	1 (1%)	40 (40%)	5 (5%)	100
UNUSED	3 (3%)	60 (58%)	1 (1%)	34 (33%)	5 (5%)	103
INDEF-NEW	--	37 (64%)	--	17 (29%)	4 (7%)	58

Table 1: Accents on *single*-word referring expressions (pronouns, proper names, bare nouns), depending on information status

Information Status	No Accent	Rise (L*H)	Rise cont. (L*)	Fall (H*L)	Fall cont. (H*)	100%
GIVEN-NONPRONOUN	28 (12%)	118 (51%)	4 (2%)	57 (25%)	25 (11%)	232
BRIDGING	16 (7%)	114 (49%)	--	81 (35%)	20 (9%)	231
UNUSED	9 (2%)	238 (48%)	1 (0%)	232 (47%)	16 (3%)	496
INDEF-NEW	--	73 (36%)	1 (0%)	121 (59%)	9 (4%)	204

Table 2: Accents on *two*-word referring expressions (det-noun, prep-noun, title-name, compound), depending on information status

References K. Eckart et al. *An Infrastructure for More Reliable Corpus Analysis*. Workshop on Web Services and Processing Pipelines in HLT. LREC, Malta, 2010 • J. Mayer. *Transcription of German Intonation. The Stuttgart System*. ms, 1995 • A. Riester et al. *A Recursive Annotation Scheme for Referential Information Status*. LREC, Malta, 2010