Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Carnegie Mellon University
Language Technologies Institute

AN EMPIRICAL ANALYSIS OF THE CORRELATION OF SYNTAX AND PROSODY

ARNE KÖHN, TIMO BAUMANN, OSKAR DÖRFLER

MAIN IDEA

Are words spoken differently depending on their syntactic function?

E.g., **do subjects prosodically differ from objects?** (significantly and all other aspects being equal).

We examine **pitch**, **power**, **duration** of a word, and the length of the **pause** after it. → low-level and objective

We model the acoustic properties using textual features, **with/without syntactic function**.
Significantly better model → significant correlation

			det↗	attr↗				aux↗…		
		subj~obja	…↗(subj~obja)		s~neb	s~rel	s~aux	(s~neb)	(s~rel)	(s~aux)
pitch	p-value	***	ns	**	ns	***	ns	ns	***	***
	effect in Cent	19.38	—	25.14	—	39.38	—	—	22.81	53.77
power	p-value	***	0.06	*	*	***	ns	ns	ns	ns
	effect in dB	-0.01	0.005	0.01	0.01	0.02	—	—	—	—
duration	p-value	***	***	ns	***	***	***	0.12	***	*
	effect in ms	35.42	-14.62	—	-17.8	-23.03	-29.32	15.88	19.03	30.04
pause	p-value	*	**	ns	ns	*	***	ns	***	ns
	effect in ms	3.81	6.14	—	—	-5.52	-10.17	—	10.73	—

STATISTICAL APPROACH

likelihood ratio test: basic and extended model to fit property, check merit of extension

basic:
property ~ 1 + *textual features* + *speaker normalization*

extended:
property ~ *basic* + *syntactic function*

We then

- Compare goodness of fit of *basic* vs. *extended*
- fit(*extended*) ≫ fit(*basic*) → significance
- coefficient of syntactic function → effect size

Textual features for estimation:

- **predicted word duration**: "expected" word length as per MaryTTS (*tpred*)
- Word **position** in sentence and **sentence length**: "late" words spoken differently from "early" ones (*wpos* and *slength*)

Normalization to account for speaker variation:

- z-normalized pitch (in semitones)
- condition *tpred* on article to normalize for tempo

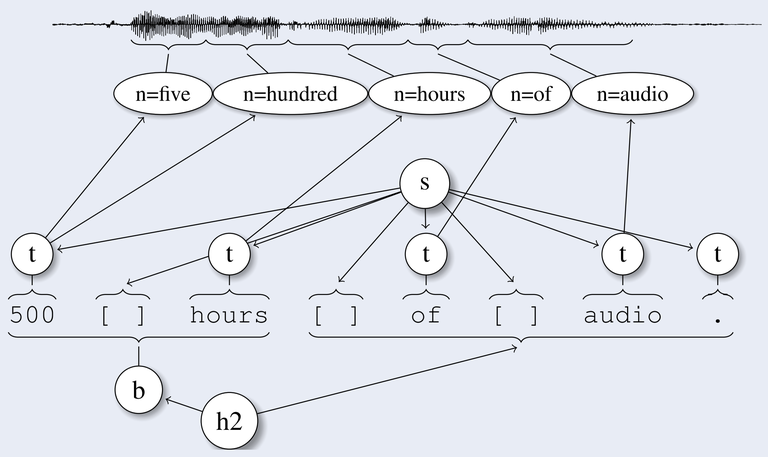
E.g. *duration* ~ 1 + *tpred* ∗ *wpos* ∗ *slength* + (*tpred*|*article*)

DATA SOURCE

Spoken Wikipedia Corpus for German:

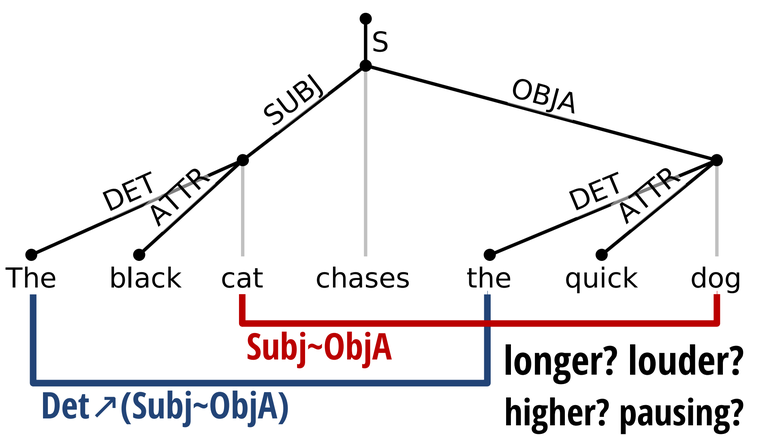
- 46 h of segmentally aligned speech
- 31,803 sentences
- 348,062 word tokens

(out of 386 h of total audio in >1000 spoken articles).



Parsed with TurboParser trained on the **Hamburg Dependency Treebank**. Audio features extracted with `SNACK`.

EXAMPLE



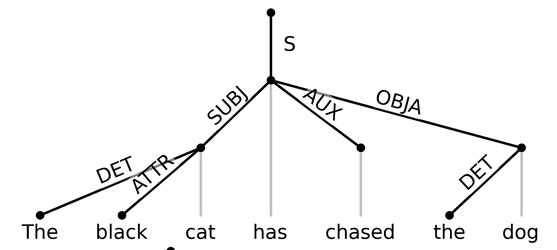
We compare words that can occur in two syntactic funtions (e.g., **Subj~Obj**).

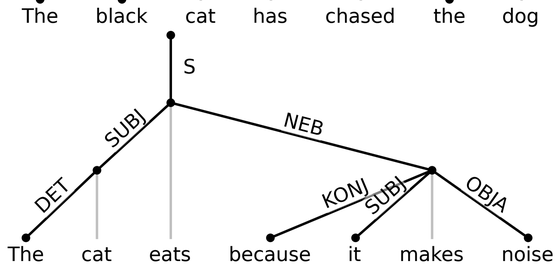
We also compare words with the same function but attached to words with different functions (e.g., **Det↗(Subj~ObjA)**).

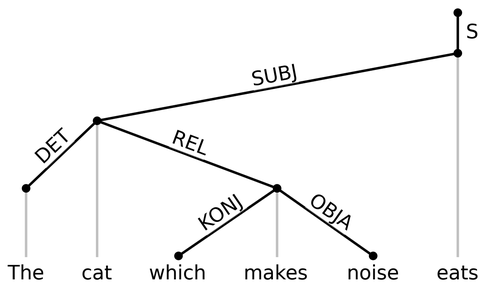
We aggregate over all sentences, i.e., different syntactic functions need not occur in the same sentence (as in the example).

Examples for **verb function**:

s: verb or auxiliary that is the head of a sentence
aux: full verb attached to an auxiliary
neb: verb that heads a subordinate phrase
rel: verb that heads a relative clause







ANALYSIS AND DISCUSSION

Highly significant effects!

Effect sizes often well above *just noticeable difference* for pitch and tempo; other effects might subconciously help in disambiguation.

Regarding nouns and words modifying nouns:

Subjects are longer and higher pitched than objects (accusative objects).

Signal power is slightly lower for subjects – maybe because energy is spread out further (longer duration)?

Determiners of subjects are significantly shorter – partially making up for the longer subjects?

Attributive adjectives of subjects are higher pitched, even more than the subject itself.

Regarding verbs:

All of *neb*, *rel*, and *aux* are spoken substantially longer than main verbs.

Verbs heading **subordinate phrases are most similar to full sentences**; the structural similarity is mirrored in prosody.

Relative clauses are **interjected phrases** that modify nouns. Their verbs are spoken with **lower pitch, less power and lengthened**, i.e., less pronounced, possibly to distinguish the additional information from the main content of the sentence.

ARNE KÖHN, ARNE@CHARK.EU
Natural Language Systems Group
Universität Hamburg

TIMO BAUMANN, MAIL@TIMOBAUMANN.DE
Language Technologies Institute
Carnegie Mellon University

https://nats.gitlab.io/swc/