Modelling Register Variation in HPSG
Stefan Müller & Antonio Machicao y Priemer
HU Berlin, Institute for German Language and Linguistics, Syntax Lab

1. Preliminary assumptions

- **Grammatical competence** (GC): Knowledge about how to combine language specific elements (word, phrases, etc.) to build form-meaning pairs (e.g. sentence)
- **Register competence** (RC): Knowledge about in which situations form-meaning pairs can be used and how in these situations form maps to meaning (e.g. impression, metaphors, irony, etc.) → possibility
- GC and RC are not (always) acquired at the same time:
  - In many cases, children know how to build a grammatical utterance, but they do not know (yet) when to use it.
  - RC is related to the acquisition of social skills, but it is closely intertwined with GC at all levels of grammatical description (cf. (1)–(4)).

(1) Phonology with GC:
   a. ich habe ...
   b. ich hab ...
   'I have ...'

(2) Lexical selection:
   a. Herr XY
   b. Digga

(3) Syntax:
   a. Das Kind hat aber nicht mal geweint!
   b. Hat aber nicht mal geweint das Kind!
   'The child has not even cried.'

(4) Semantics:
   a. I arrived at 3:35 pm.
   b. I arrived at 4.

Furthermore, RC is related to extra grammatical information (5), to phenomena combining different grammatical levels (6).

(5) Speakers intentions (pragmatics) of GC:
   a. e.g. intentionally violating Grice’s maxims to convey information
   b. e.g. using a dialect to convey familiarity

(6) Interface: Phonology–Morphology–Syntax–Information Structure:
   a. Wir haben das gekauft.
   b. Hamwa gekauft.
   'We have bought it.'

2. Which kind of model do we aim for?

We use the HPSG formalism since . . .

- it allows us to combine different levels of linguistic description
- it can integrate extra-grammatical levels of language description to account for the different phenomena related to register variation
- it can be implemented as a computer processable grammatical model (cf. Müller 2014, 2015)
- it is a performance-compatible competence grammar (cf. Sag & Wasow 2011; contra Newmeyer 2003; Adger 2006)
- it can be expanded by probabilistic methods (cf. Bow 1995; Abeysinghe 1997; Miyao et al. 2008)
- it can be expanded to domains bigger than a sentence (e.g. texts)
- The model must be able to predict how a probabilistic sign (e.g. word, sentence, text) is in a specific register (register comparison).
- The formalism has to allow the combination of register values of lexemes, phonological or syntactic constraints, etc. to compute the register value of the linguistic sign at different levels of complexity.

3. Multiple grammars

Following the conceptualisation in multilingual grammar engineering (cf. Bender et al. 2002; Müller 2015), one possible way to model registers is assuming a single set of constraints (viz. a grammar) for each register.

Here, we follow the idea of the CoreGram project (Müller 2015) where grammars are organised as sets of constraints. Assuming only two grammars (German and Dutch), we will find shared constraints (e.g. w.r.t. argument structure, V2, SOV, etc.). These constraints are organised in a set (Set 3) that is shared by the grammars of both languages. Adding a further language (e.g. Danish), we can organise the constraints shared between Danish and the other two languages in a more general set (Set 5).

In the same way, registers can be organised as sets containing the constraints for the use of signs in specific registers. Building on work from the 1st phase of the SFB (cf. Schäfer et al. 2002), we assume different probability distributions for linguistic signs in different registers. That is, the GC is constrained in Set 3, but the probabilities for the constraints to be realised in one register or the other are organised in different sets. The probabilities are represented by w* values that can be assigned to affixes, words, constraints, etc. and by means of register constraints can be computed together reflecting the probability of an utterance in a register.

4. Conclusions and open questions

- Cooperation with other projects in order to identify register related phenomena and provide adequate formal analyses for them.
- Discussions about possible models and what they have to perform are needed.
- Mathematics behind probabilistic HPSG has to be worked out in detail.
- Assuming a specific situation and knowledge about the register for such situation: How can we recognise register violations and what do they mean (change of grammar vs. inappropriate use)?

(7) Mit Verlaub, Herr Präsident, Sie sind ein Arschloch.
   [Joscha Fischer, 1984]
   ‘With respect, Mr. President, you are an asshole.’

References


