Institut für Linguistik
Institut für Maschinelles Sprachverarbeitung

Artemis Alexiadou
Point of departure for the SFB 732: pervasive ambiguity that characterizes all levels of linguistic analysis.

Goal: to achieve a better understanding of and model the mechanisms that lead to ‘ambiguity’ control/disambiguation.

Hypothesis: We regard ambiguity as the result of underspecification and hence disambiguation processes as processes of specification of an underspecified input in context.

Reviewers' complaint: how do you understand underspecification?
'Der Begriff der Unterspezifikation wirft Fragen hinsichtlich seiner Konzeption auf. Eine größere Kohärenz in der konzeptuellen Interpretation des Begriffs und der Bezug zu Ambiguität sowie insbesondere eine genauere und explizite Darstellung dieser Problematik werden für die nächste Förderperiode erwartet'.
Before we talk about the future then, let us take a step back and clarify the issue.

In the first longer part of this presentation, I will discuss basically two usages of the term underspecification that involve ‘compact‘ representations vs. competition between forms with varying degrees of specificity (elsewhere condition): semantic vs. morphological underspecification.

Other understandings of the term are presumably relevant/necessary to clarify the issue and achieve coherence.

In the end, I will offer some remarks about phase III and try to define the term less well-behaved data.
Underspecification

- In the case of an ambiguity, the number of alternative readings can be large, but is finite.

- **Economy principle:** ambiguity may not/does not depend on there being a plurality of representations but on there being an **underspecified** representation.

- Plurality of representations is un-economical as it leads to a series of problems (the construction of representations for all the possible readings of a given sentence/word is a computationally extremely expensive task).

- The construction and use of compact representations may allow a much more efficient way of processing.
Underspecification

- Other arguments in favor of underspecification that have been discussed in the literature involve:
  
  - i) machine translation, since ambiguities in a sentence in the source language are often retained in the target language.
  
  - ii) psycholinguistic considerations.
  
  - ...

Definition

- Underspecification emerged in phonology and was later adopted by semanticists to model ambiguity.

- Underspecification can be broadly defined as the deliberate omission of information from linguistic descriptions.
Semantic underspecification

- In **semantic underspecification**, the goal is to capture several alternative realisations of a linguistic phenomenon in one single representation.

- This avoids a disjunctive enumeration of alternatives.
- The ambiguities covered by semantic underspecification are structural.
- Structural ambiguity arises if one (morpho-)syntactic structure corresponds to several meanings.
Fields

- Fields of application:
  - scope
  - ellipsis
  - lexical ambiguity
  - metonymy
  - aspectual coercion
  - incompletely transmitted messages
  - ...
Formalisms

- Underspecified representation formalisms must obey certain requirements (see e.g. Ebert 2005 for an overview and summary).
- These formalisms include
  - Quasi-Logical Form (QLF; Alshawi 1992)
  - Underspecified Discourse Representation Theory (UDRT; Reyle 1993)
  - Hole Semantics (HS; Bos 2004)
  - Glue Language Semantics (‘Glue’: Dalrymple et al. 1997)
  - Constraint Language for Lambda Structures (CLLS; Egg et al. 2001)
  - Minimal Recursion Semantics (MRS; Copestake et al. 2005)
  - Regular Tree Grammars (Koller et al. 2008)
- .....
Underspecification makes possible very flexible syntax-semantics interfaces.

There is no need to rearrange syntactic structures for semantic processing.

Immediately compatible with surface-oriented syntactic frameworks which can be used as the basis of semantic modules with underspecification: e.g. LFG, HPSG, etc.

Various benefits for a large number of NLP applications.
Our hypothesis had two further aspects: if ambiguity resolution is a process of specification, then

- a) this transforms one linguistic representation into any one of several alternative more specific representations,
- b) in doing so, a choice has to be made between these on the basis of the evidence taken from some relevant context.

- Context provides constraints and triggering conditions;
- Relationships between representations emerge.
(1) Linguistic expressions (LEs) may lack some property x at some level of representation, and exhibit the same property x at some other level of representation.

- The relationship between the two levels of representation is determined by various rules or principles, which can have various forms depending on the theoretical framework.
- Levels need not be adjacent (in the strict understanding of the pipeline model). We will come back to that.
Model of specification:

Pipeline vs. Joint model.

Importantly: Principle of Economy

Rather than assuming an explicit listing of the entire set of choices \( \{a_1 \ldots a_n\} \) prior to contextual disambiguation, assume a compact representation.
Syntax-morphology interface

- In languages like English, and German, passives and reflexives are distinct.
- In a number of other languages, the same form is used to express reflexive and passive syntax: **syncretism**. (1) is ambiguous between the two readings.

(1) **O Janis plithike**

    John washed-non-active-3sg

John washed himself/John was washed by somebody
Syntax-morphology interface

- **Underspecification** (Embick 2004): the Nact affix of Greek contributes only one of the shared requirement of the English constructions (absence of an overt external argument).

- However, the syntax and the semantics of the passive vs. the syntax and the semantics of the reflexive are distinct.
Syntax-morphology interface

- Other cases in point: nominalizations; participle formation

(2) Italian examples, the phenomenon is identical in German, English, etc.:  
(EVENT)
a. Per evitare l'ostruzione del tubo i tubi stessi devono essere lavati.  
‘To prevent the obstruction of the pipes, pipes must be cleaned’  
(STATE)
b. L'ostruzione può essere temporanea o permanente.  
‘The obstruction may be temporary or permanent’  
(RESULT-OBJECT)
c. Questo test permette di capire esattamente dove si trova l'ostruzione.  
This test allows to understand exactly where the obstruction is’  
(3)  
a. The book was written passive  
b. The book is well-written adjectival
Syntax-morphology interface

- In these two examples, the syntactic and semantic features associated with the form *written* and *ostruzione* are very different.
- In all cases and interpretations the affixes–*en* and –*zione* are used.

- The affixes –*en* and –*zione* are **underspecified** for the syntactic structures in which they occur.
- The structures themselves, i.e. the syntactic representations, however, are fully specified.
- There is a shared requirement for all readings of the nominalization and participle that these affixes satisfy.
Syntax-morphology interface

- Dealing with linguistic phenomena in morphology by appealing to Underspecification makes reference to the Elsewhere condition; this means that we determine the choice on the basis of specificity,
- and we appeal to the notion of default/general.
- A default/general/underspecified form lacks a set of features, and because of that it is thus compatible with a wide range of potential environments.
Syntax-morphology interface

- NB. For at least some of the cases we were looking at, this view on underspecification is tied with assumptions that morphology is realizational, for some the null-hypothesis.

- Realizational theories of morphology assume that the root's association with a particular set of morpho-syntactic properties licenses the introduction of inflectional exponents.

- For example, in *likes* the association of the root *like* with present tense, third person agreement, and indicative mood license the attachment of –s.
Syntax-morphology interface

- Underspecification of exponents gives rise to competition (because it is often the case that more than one exponent fits into a syntactically defined, fully specified context). The competition can be resolved by a constraint like the Subset Principle in (4).

(4) *Subset Principle*: A vocabulary item V is inserted into a functional morpheme M iff (i) and (ii) hold:

- (i) The morpho-syntactic features of V are a subset of the morpho-syntactic features of M.
- (ii) V is the most specific vocabulary item that satisfies (i).

- NB: local competition vs. avoidance of enumeration of alternatives.
Syntax-morphology interface

- The formulation of the Subset Principle in (4) presupposes late insertion of inflectional exponents (‘vocabulary items’) into abstract syntactic terminal nodes (‘functional morphemes’), as it is envisaged in Distributed Morphology.

- The Subset Principle ensures that an exponent can only be used in a given syntactic context if it is compatible with this context; i.e., the (possibly underspecified) exponent must be characterized by a subset of the morpho-syntactic features of the syntactic context. Furthermore, out of the set of exponents that satisfy the compatibility criterion, the exponent is selected that is more specific than its competitors.
Syntax-morphology interface

Specificity of exponents can be defined as in (5), e.g. Müller (2005).

(5)  **Specificity of vocabulary items:**
A vocabulary item $V_i$ is more specific than a vocabulary item $V_j$ iff there is a class of features $F$ such that (i) and (ii) hold.

(i) $V_i$ bears more features belonging to $F$ than $V_j$ does.
(ii) There is no higher-ranked class of features $F'$ such that $V_i$ and $V_j$ have a different number of features in $F'$.

In e.g. Greek Voice, there is no more specific exponent; hence NAct is used.
Summary:

- the most desirable way to treat syncretism is via underspecification and this gives rise to elsewhere forms.
- Underspecification of the type understood within Distributed Morphology involves individual vocabulary items, and treats syncretism as a lexical phenomenon.
- The syntax/morphology interface must be designed so as to ensure that underspecified representations in morphology correspond to fully specified representations in syntax.

The question arises of why morphology and syntax should differ in this respect.
(partial) Synthesis

- Two notions?

- Broad definition:
  - Underspecification = omission of information from linguistic descriptions
  - Semantic underspecification (underspecification I):
    - one (morpho-)syntactic structure corresponds to several meanings
  - Morphological underspecification (underspecification II):
    - one element satisfies a shared requirement of at least two different fully specified syntactic representations.

- Moreover: no disambiguation at all.
Teasing the two notions apart

Are there test cases from the area of morpho-syntax which would enable us to tease the two alternatives apart?

We believe that indeed Voice syncretisms in Greek are:

(1) O Janis plithike \(\text{passive/reflexive}\)
    John \(\text{washed-non-active-3sg}\)
    John washed himself/John was washed by somebody
Teasing the two notions apart

- As mentioned, traditionally it has been assumed that it involves a default form (NAct is the general form used in all transitivity alternations.)
- But: what if the above ambiguity involves underspecification type I?
- Then there is no formal difference between a reflexive and a passive. In languages of this type, the passive will systematically show the same allomorphy with the reflexive, simply because these *will* be identical.
Other uses of the term underspecification:

- lexical underspecification:
  - (i) roots stored in the lexicon without specification of category.
  - (ii) *grain-size* for defining verb classes, useful for understanding member behavior and overall lexical organization:
    
    *word's lexicalized meaning* vs. facets of meaning attributable to context (Levin and Rappaport Hovav 2011): within particular semantic domains there can be verbs that describe bringing about *results* and others that describe carrying out activities—*manners* of doing.

- ......
The future

- In phase I of the SFB we have been focussing on determining the appropriate representations of the various candidates to be computed.
- In the second phase of the SFB, we focused on modelling the interaction effects between two neighbouring components and dealing with well-behaved, standard data.
- In phase III, we will be concerned with broader effects across distant levels and sub-processes of specification and comparing parallel subprocesses.
- The empirical domain will be extended to what we will call non-standard data in a way we will try to make precise.
The future

- Examples of the interaction of levels that are not-adjacent.

- Some examples will be discussed later today.

- In e.g. area B: Interaction between morpho-syntactic structure and linguistic ontology/conceptual structure.

- In our Voice example: one syntactic representation is disambiguated (maybe) at the level of conceptual structure.
- Standard/well-behaved vs. non-standard/less well-behaved
- Written vs. oral
- Synchronic vs. diachronic
- Well-studied vs. less studied
- Core vs. peripheral
- Regular vs. special
- L1 vs. L2
- ....

These will help us test our models/theories.