

Applying Constraints derived from the Context in the process of Incremental Sortal Specification of German *ung*-Nominalizations

Kristina Spranger and Ulrich Heid

IMS, Universität Stuttgart, 70174 Stuttgart, Germany

Abstract. Many German nominalizations with the affix *-ung* are sortally ambiguous. Within a sentence, lexico-semantic and/or syntactic phenomena may support disambiguation. The sortal interpretation of a nominalization may vary depending on the underlying syntactic analysis of one and the same, syntactically ambiguous sentence.

We model the process of sortal disambiguation as a constraint-based incremental process. The process is incremental as it evaluates in subsequent steps constraints from increasingly larger context windows.

1 Introduction

In this article, we present work towards the automatic interpretation of German nominalizations with the affix *-ung*, such as *Lieferung* (*delivery*) or *Messung* (*measurement*). Many such *-ung*-nominalizations are ambiguous with respect to their sortal interpretation (cf. [3] - who lean heavily on [8] and [7] - for the notion of sortal ambiguity). In section 2, a more detailed discussion on sortal ambiguity as regards German *-ung*-nominalizations is given.

We are working towards a system for data extraction from corpus text that is able to carry out sortal disambiguation. Given the productivity of the *-ung*-formation process in German (cf. [4] and [12]) and the high frequency of *-ung*-nominalizations in text (cf. [6] or [10]), this ability is relevant, among others, for question answering or high quality information extraction¹.

In this work, we analyze the influence of the context of an *ung*-nominalization on its sortal interpretation. Relevant contextual phenomena include lexical combination partners of the nominalization and/or the surrounding syntactic structures. As we rely on parsed corpus data, we have no discourse representations available that go beyond the sentence level. Thus, we have to limit the interpretation process to the sentence context, even though the disambiguation of some *-ung*-nominalizations would require a larger context.

In a preliminary case study (see section 3), we have identified some of the contextual phenomena which constrain the sortal interpretation of *Messung*. From a

¹ A more detailed discussion on the relevance of this ability in natural language processing systems is given in [11].

descriptive perspective, such phenomena serve as “indicators” of sortal readings. We model the process of sortal interpretation as a process of incremental specification where the context of a nominalization is used for its sortal interpretation. Knowledge about the reading indicators is explicitly formulated as constraints that are applied to a given nominalization. In section 4, we explain the specification process in detail.

In the process of sortal disambiguation, the order in which different constraints are applied is crucial for the interpretation of a nominalization at the sentence level. Obviously, the order in which the constraints are applied depends on the syntactic analysis of the targeted sentence, and syntactic ambiguity leads to multiple syntactic analyses. In section 5, we demonstrate that the sortal interpretation depends on the underlying syntactic analysis.

We conclude in section 6, addressing relevant aspects of a planned underspecified representation to make allowance for the effect of syntactic ambiguity and pointing to some more directions of future work.

2 Ambiguous German Nominalizations with *-ung*

German verb nominalizations with *-ung* are up to three-fold ambiguous concerning their sortal interpretation. They may have an event reading, a (result) state reading, and an object reading².

2.1 The Sortal Interpretation of *ung*-Nominalizations

According to [3] the primary distinction is the distinction between (i) eventualities and (ii) objects:

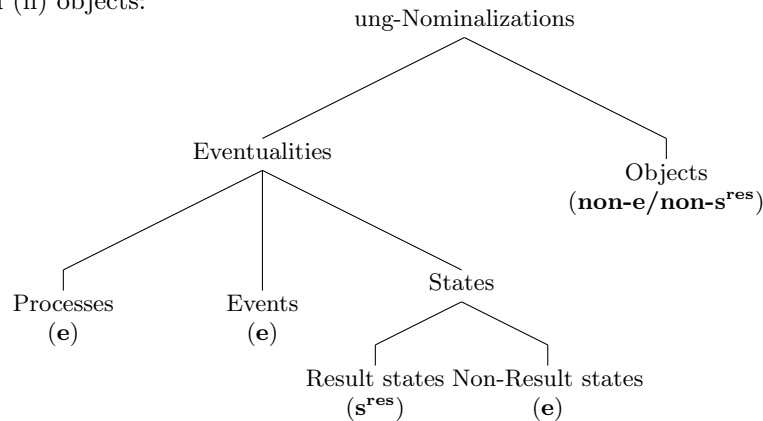


Fig. 1. The Sortal Interpretation of German *-ung*-Nominalizations

² For a more detailed discussion cf. [3], and [13] and [5] whose works are based on the theory developed in [3].

Eventualities Ehrich and Rapp subsume processes, events, and states under the concept of eventualities taken over from [2].

Events refer to telic actions whereas processes refer to atelic actions³. According to [9] processes as well as events can be seen as event complexes that are an association of a goal event, or “culmination” with a “preparatory phase” by which it is accomplished and a “consequent state” which ensues.

States (result states as well as non-result states) refer to eventualities that do not have a dynamic preparatory phase. Result states (e.g. *Absperrung (roadblock)*), in contrast to non-result states (e.g. *Bewunderung (admiration)*), are caused by a preceding event. Therefore, we distinguish between result states and other eventualities (including non-result states).

In the following, processes, events and non-result states are referred to by **e**, result states are referred to by **s^{res}**.

Objects Objects refer to physical as well as abstract objects. They are referred to by **non-e/non-s^{res}**.

2.2 Distributional Tests

Except for non-result states and objects all classes of *-ung*-nominalizations (cf. figure 1) refer to some phase in the event complex as it is described by Moens and Steedman (cf. [9]): result states refer to the post-culmination phase, and events and processes refer to the whole event complex. Thus, it is especially challenging to keep them apart. To this end, Ehrich and Rapp propose a number of distributional tests:

1. Only eventualities allow to refer to phases of the events (a) and can be combined with process modifying predicates (b):

(a) Die Verfolgung des Täters / Die Absperrung des Geländes
The pursuit the perpetrator / The cordon the area

beginnt / hört auf / wird unterbrochen.

starts / stops / is interrupted.

‘The pursuit of the perpetrator / The cordon of the area starts / stops / is interrupted.’

(b) die **umständliche / vorsichtige** Verfolgung des Täters /
the awkward / cautious pursuit the perpetrator /

Absperrung des Geländes

cordon the area

‘the awkward / cautious pursuit of the perpetrator / cordon of the area.’

2. Result states can be combined with stative predicates (a) and with predicates of perceptibility (b) (summed up as “static predicates”):

³ According to [15] events are “accomplishments” and “achievements”, and processes are “activities”.

- (a) die **bestehende** Absperrung des Geländes
the existing cordon the area
 'the existing cordon of the area'
 - (b) die **vorgefundene / kartographisch registrierte** Absperrung
the found / cartographically registered cordon
 des Geländes
the area
 'the cordon of the area found / cartographically registered'
3. Duration predicates can only occur together with processes and result states:
- die **tagelange** Verfolgung des Täters / Absperrung des
the lasting for days pursuit the perpetrator / cordon the
 Geländes
area
 'the pursuit of the perpetrator / cordon of the area lasting for days'
4. Events can go together with time frame predicates (a) and they allow to refer to the incremental progression of the event (b):
- (a) die **in zwei Tagen** erfolgte Absperrung des Geländes
the in two days accomplished cordon the area
 'the cordon of the area accomplished in two days'
 - (b) die **allmähliche** Absperrung des Geländes
the gradual cordon the area
 'the cordon of the area completed step by step'

The distributional tests show that event nominalizations and result state nominalizations are distributed complementarily.

3 The *-ung*-Nominalization *Messung*: A Case Study

The nominalization *Messung* (*measurement*) is two-fold ambiguous: it allows for an event interpretation (e), and for an object interpretation (**non-e**)⁴.

3.1 Sortal Readings of *Messung*

The event reading of *Messung* refers to the process of measuring. Sentence (1) is a typical context for *Messung* as an event.

- (1) die **Messung** des Erdumfangs durch Eratosthenes
the measuring the circumference of the earth by Eratosthenes
 'the measuring of the circumference of the earth by Eratosthenes'

The object reading refers to the result of a measuring process, i.e. to data or figures. Sentence (2) is a context for *Messung* as an object.

⁴ For the sake of convenience, we do without **non-s**^{res} since there is no result state interpretation of *Messung*.

- (2) Die **Messungen** liegen unter dem zulässigen Grenzwert von 250 ppm.
The measurements lie under the acceptable critical value of 250 ppm.
 ‘The measurements are lower than the maximum permissible value of 250 ppm.’

3.2 Disambiguating Reading Indicators from the Context

To decide about the sortal interpretation of an *-ung*-nominalization, humans seem to use lexico-semantic and syntactic reading indicators from the context. Many lexical indicators are combinatory constraints of lexico-semantic nature, ranging from preferences for general (ontological) classes, over selection restrictions, to lexeme-specific combinations. Some such indicators have been used by [3] to formulate their distributional tests (cf. section 2.2). We list more such indicators for event and object readings of *Messung* in tables 1 and 2. These indicators have been derived from a manual analysis of circa 400 sentences newspaper text.

Type	Examples
Reference to Event Phase	nominalization as subject: <i>Messung geht weiter</i> nominalization as object: <i>Messung aufnehmen, fortsetzen, abschliessen</i>
Duration predicates	adjectives: <i>fortlaufende, kontinuierliche Messungen</i> temporal PPs: <i>während der Messung</i>
Selection Restriction of Verbs of Order	<i>Messung anordnen, vorschreiben, veranlassen</i>
Lexical Collocations	support verbs: <i>Messung findet statt, Messungen durchführen</i>
Local/Temporal Adjuncts	<i>Messungen an Strassen, Messungen im Sommer</i>

Table 1. Event Indicators

Type	Examples
Static Predicates	<i>Messungen liegen vor</i>
Value Indicating Verbs	<i>M. liegt bei <value></i>
Use with Proving Verbs	<i>Messung beweist/zeigt, dass; jmd. zieht aus der Messung den Schluss, dass</i>

Table 2. Object Indicators

In a given sentence, the lexical indicators may appear in different syntactic structures. For example, a support verb which has the nominalization as its object may also come as a prenominal participle or in a relative clause. Moreover, roughly synonymous indicators may belong to different word classes.

4 Incremental Sortal Specification in Context Using Context-Derived Constraints

Taking the contextual phenomena we have identified to constrain the sortal interpretation of *Messung* as a starting point, we will show in the following how the knowledge about these reading indicators is explicitly formulated as constraints, and how these constraints are used in an incremental process of sortal disambiguation.

4.1 Competing Reading Indicators

In many cases there is more than one indicator in a sentence, and not all indicators present in a given sentence support the same sortal reading. Sentence (3), for example, contains two indicators: one for the object reading and one for the event reading.

- (3) Die Geologen beschreiben Messungen [auf den Seychellen]_e, [die *The geologists describe measurements on the Seychelles, that Anzeichen des Klimawandels zeigen*]_{non-e}.
indications the climate change show.

'The geologists describe measurements on the Seychelles that show indications of the climate change.'

auf den Seychellen is an indicator for the event reading: it is a local adjunct (cf. table 1). The relative clause *die Anzeichen des Klimawandels zeigen* with *zeigen* as predicate is an indicator for the object reading: *zeigen* belongs to the class of "proving verbs" (cf. table 2).

Nevertheless, the nominalization does not (necessarily) remain sortally ambiguous at the sentence level. The human reader is perfectly able to interpret the nominalization as an event or as an object - at the latest when he considers a larger context window than one sentence. Obviously, there are cases where human readers are not able to disambiguate the sortal interpretation of the nominalization (cf. sentence 4). However, these seem to be cases where it is not relevant for the comprehension of the text.

- (4) Die Schifffahrt profitiert von den aktuellen Messungen über
The navigation benefits from the current measurements over
Windgeschwindigkeit und Wellenhöhen.
wind speed and wave heights.

'The navigation benefits from the current measurements of wind speed and wave heights.'

On the other hand, there are examples that show that some contextual phenomena enforce a certain interpretation which leads to a misinterpretation of a sentence such as sentence 5.

- (5) Da die Messungen vor den Fenstern der vom Lärm
Since the measurements in front of the windows the of the noise
Betroffenen vorgenommen wurden, hat ihnen die Gemeinde
persons concerned carried out were, has them the municipality
Schallschutzfenster angeboten.
soundproof windows offered.

'Since the measurements were carried out in front of the windows of the persons concerned of the noise, the municipality offered them soundproof windows.'

4.2 The Incremental Process of Sortal Specification

What determines the sortal interpretation at the sentence level is the underlying syntactic reading. Due to syntactic ambiguity we get most often more than one syntactic reading. Different syntactic readings may lead to different sortal interpretations of one and the same sentence.

The sortal interpretation depends on the syntactic reading insofar as the indicators of a given sentence may appear in different places of the according syntax tree. Since indicators of different sortal readings may be present in one and the same sentence, we assume that the interpretation process works strictly incrementally, i.e. the indicators "enter" the context used for interpretation one after the other. Depending on what context window is considered the sortal interpretation may vary; if the context window grows the sortal interpretation may change.

The interpretation that is accessible for a context larger than one sentence is the one at the sentence level. So, in order to come up with the sortal interpretation of the nominalization at the sentence level we start with the nominalization in the null context and walk up the syntax tree. The considered context grows step by step walking through the syntax tree.

The sortal interpretation of the nominalization is "defeasible" as long as there is a larger context that is relevant for the interpretation process; in its current context, the sortal interpretation is "indefeasible". As regards our concept of "defeasible" and "indefeasible" sortal interpretations, we lean on Alshawi and Crouch's concept of "believed" vs. "unbelieved" in their monotonic semantic interpretation (cf. [1]). The sortal interpretation of a nominalization is explicitly called into question until there is no larger context relevant for the interpretation process. That means, the sortal interpretation of a nominalization actually is a disjunction of all possible sortal types with the "contextually active" sortal type "underlined". Therefore, the interpretation process is a monotonic process nevertheless.

The core idea of this specification process is that the reading indicators that enter the context while it grows incrementally introduce constraints that can be applied to a nominalization in its current context. The specification process follows the algorithm given below:

- The “bare” *-ung*-nominalization (i.e. the nominalization in its null context) which, obviously, is sortally ambiguous gets the sortal type $\langle \mathbf{e} \dot{\cup} \mathbf{non-e} \rangle^5$.
- Then, all sibling nodes are considered: before a sibling node is added to the “active” context⁶, it is checked whether it dominates an indicator.
- If so, the indicator introduces a constraint over the interpretation of the *-ung*-nominalization in its current context.
- The constraint is applied, and the sibling node is added to the context of the nominalization.
- The procedure is repeated until the sentence node is reached.

The Main Constraint and a Type Conversion Function Supposed:

- $U = \{x, x \text{ is a } -ung\text{-nominalization}\}$
- $m \in U$
- $\mathbf{ung-sort} = \{e, s^{res}, \mathbf{non-e}/s^{res}\}$
- $\alpha, \beta \in \mathbf{ung-sort}$

We define a constraint $C_{\langle \alpha \dot{\cup} \beta, \alpha \rangle}$ that has the following two properties:

1. $C_{\langle \alpha \dot{\cup} \beta, \alpha \rangle} (m_{\alpha \dot{\cup} \beta}) = m_\alpha$
2. $C_{\langle \alpha \dot{\cup} \beta, \alpha \rangle} (m_\beta) = m_\alpha$

In order that the constraint be applicable to m_β we define a type conversion function τ :

- $\tau(m_\beta) = m_{\alpha \dot{\cup} \beta}$

5 Four Syntactic Readings and their Corresponding Sortal Readings

In order to demonstrate how the specification algorithm works, we will analyze sentence (3): *Die Geologen beschreiben Messungen auf den Seychellen, die Anzeichen des Klimawandels zeigen*.

5.1 Four Syntactic Readings

Sentence (3) has at least four syntactic readings since there are two possible attachment points for the prepositional phrase *auf den Seychellen* and two for the relative clause *die Anzeichen des Klimawandels zeigen*. These four most obvious syntactic analyses are given below (our syntax trees exclusively reflect the dominance relations and we abstract away from linear surface order):

1. *Die Geologen* is the subject, *Messungen* is the direct object of *beschreiben*, the relative clause is attached to *den Seychellen*, and the emerging complex prepositional phrase is an adjunct of *beschreiben* (cf. figure 2).

⁵ $\langle \mathbf{e} \dot{\cup} \mathbf{non-e} \rangle$ reads event or object.

⁶ “Active” context is used in the sense of “active” edges in chart parsing.

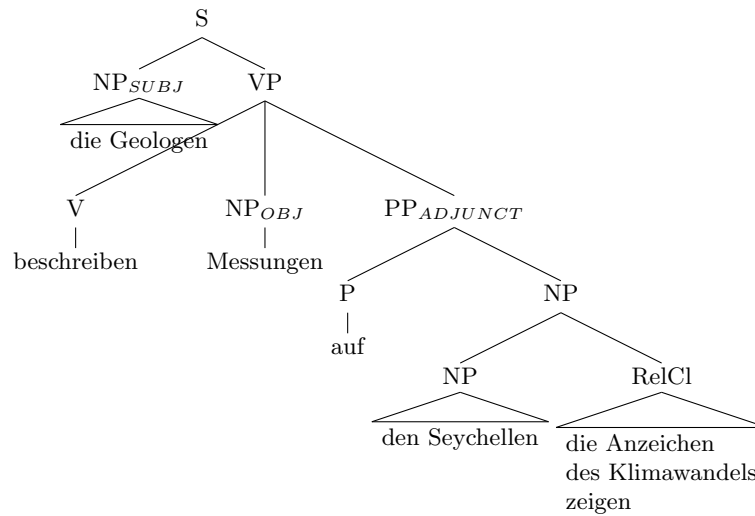


Fig. 2. Reading (1) of Sentence (3)

2. *Die Geologen* is the subject, *Messungen* is modified by the relative clause, and the resulting complex noun phrase is the direct object of *beschreiben*. The prepositional phrase is an adjunct of *beschreiben* (cf. figure 3).
3. *Messungen* is modified by the prepositional phrase. The emerging complex noun phrase is modified by the relative clause and constitutes the direct object of *beschreiben*; there is no verbal adjunct (cf. figure 4).
4. The relative clause is attached to *den Seychellen*, and the resulting complex prepositional phrase is attached to *Messung*. The emerging complex noun phrase is the direct object of *beschreiben*; there is no verbal adjunct in this syntactic analysis (cf. figure 5).

5.2 The Incremental Sortal Specification of Sentence (3)

In the following, we will show that depending on which syntactic reading we choose, we may end up with different sortal interpretations. To this end, we will analyze in detail reading (3) (cf. figure 4). In all other cases, we only present the result of the specification process.

The Reading Indicators and the Constraints they Introduce In sentence (3), we find two competing reading indicators (cf. 4.1) which introduce the following constraints⁷:

⁷ Again, for the sake of convenience we do without **non-s**^{res}.

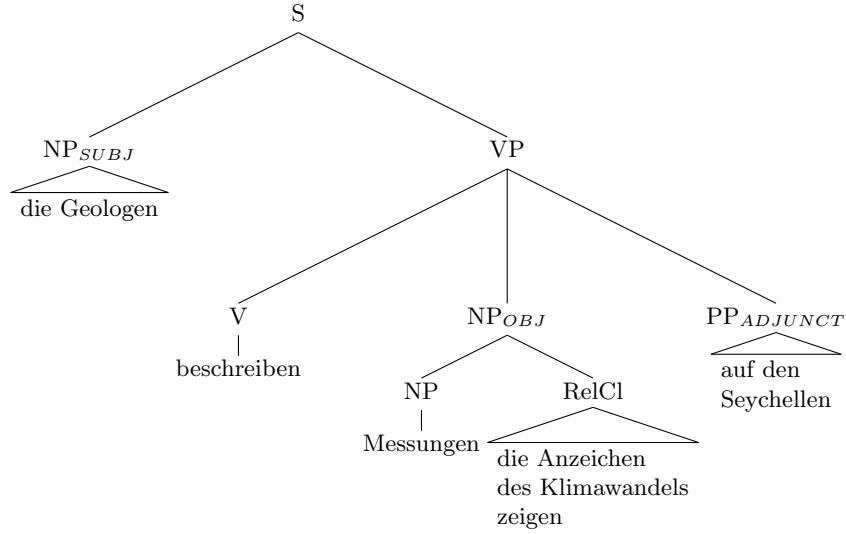


Fig. 3. Reading (2) of Sentence (3)

1. The local PP *auf den Seychellen* introduces a constraint that yields a linguistic object of the event-type $\langle \mathbf{e} \rangle$: $C_{\langle e \cup_{non-e}, e \rangle}^+$.
2. The relative clause with the predicate *zeigen* introduces a constraint that yields a linguistic object of the object-type $\langle \mathbf{non-e} \rangle$: $C_{\langle e \cup_{non-e}, non-e \rangle}^+$.

The Sortal Interpretation of Reading (3)

1. *Messungen* in the null context is considered; it is assigned the type $\langle e \cup_{non-e}, e \rangle$.
2. All sibling nodes of *Messungen* are considered: i.e. the prepositional phrase *auf den Seychellen*.
3. Does the PP-node dominate an indicator ?
Yes: The indicator introduces a constraint that yields a linguistic object of the event-type: $C_{\langle e \cup_{non-e}, e \rangle}^+$.
4. The constraint is applied to *Messungen*, the active context grows, and *Messungen* in its “new” active context is of the type $\langle \mathbf{e} \rangle$: $[\textit{Messungen auf den Seychellen}]_{\langle \mathbf{e} \rangle}$.
5. All sibling nodes of *Messungen auf den Seychellen* are considered: i.e. the relative clause *die Anzeichen des Klimawandels zeigen*.
6. Does the RelCl-node dominate an indicator ?
Yes: The indicator introduces a constraint that yields a linguistic object of the object-type: $C_{\langle e \cup_{non-e}, non-e \rangle}^+$.

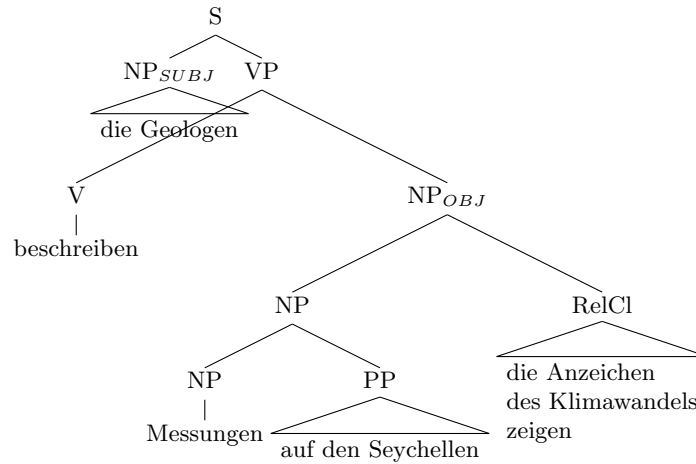


Fig. 4. Reading (3) of Sentence (3)

7. The constraint should be applied to *Messungen auf den Seychellen*, but this noun phrase is of the wrong type: it should be of the type $\langle e \dot{\cup} non - e, e \rangle$, but is of the type $\langle e \rangle$.
 8. The type conversion function τ is applied: $\tau(NP_{\langle e \rangle}) = NP_{\langle e \dot{\cup} non - e \rangle}$.
 9. Now, the constraint is applied to *Messungen auf den Seychellen*, the active context grows, and *Messungen auf den Seychellen* in its “new” active context is of the type $\langle non - e \rangle$: [*Messungen auf den Seychellen, die Anzeichen des Klimawandels zeigen*] $_{\langle non - e \rangle}$.
 10. The predicate *beschreiben* as well as the subject *die Geologen* do not introduce constraints into the context. We reach the sentence level and the interpretation process is finished.
- \Rightarrow If the underlying syntactic analysis is reading (3), the sortal interpretation is the object-interpretation.

The Sortal Interpretations of Readings (1), (2), and (4) Following the same algorithm, we end up with an ambiguous sortal interpretation of *Messungen* in case of reading (1), reading (2) leads to an event-interpretation, and reading (4) enforces an object-interpretation.

6 Conclusions and Future Work

We have shown that the sortal interpretation of *-ung*-nominalizations is highly context-dependent: the sentence context introduces indicators which can trigger a sortal reading. The interpretation is also dependent on syntactic ambiguity, as different syntactic readings of a sentence may give rise to different sortal interpretations.

We specified an algorithm for the incremental sortal specification of *-ung*-nominalizations which uses context-derived constraints in order to determine the sortal interpretation at the sentence level.

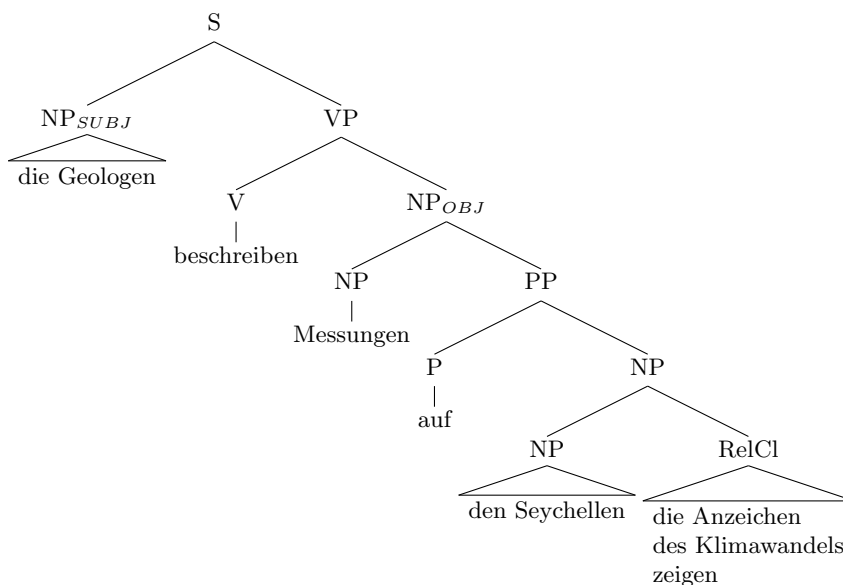


Fig. 5. Reading (4) of Sentence (3)

For an implementation in the framework of data extraction from corpus text, we will assess which syntax formalisms and parsing grammars provide adequate input for the specification process.

To be able to provide all possible syntactic readings and the pertaining sortal interpretations, we are developing an underspecified representation that should assemble all possible syntactic readings (cf. [14] for ideas that possibly can be adopted).

References

1. Alshawi, Hiyon, Crouch, Richard S.: Monotonic Semantic Interpretation. Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics (1992) 32–38
2. Bach, Emmon: The Algebra of Events. *Linguistics and Philosophy* **9(1)** (1986) 5–16
3. Ehrich, Veronika, Rapp, Irene: Sortale Bedeutung und Argumentstruktur: ungnominalisierungen im Deutschen. *Zeitschrift für Sprachwissenschaft* **19(2)** (2000) 245–303
4. Esau, Helmut: Some facts about German nominalization. *Neophilologus* **55(1)** (1971) 150–156
5. von Heusinger, Klaus: The Interface of Lexical Semantics and Conceptual Structure: Deverbal and Denominal Nominalizations. In: *Nominalisierung*. Zimmermann, I. and Lang, E. (eds.). Zentrum für Allgemeine Sprachwissenschaft, Berlin (2002) 109–124
6. Knobloch, Clemens: Zwischen Satz-Nominalisierung und Nennerivation: *-ung*-Nomina im Deutschen. *Sprachwissenschaft* **27(3)** (2002) 333–362

7. Lakoff, George: Linguistics and Natural Logic. In: Approaches to Natural Language. Davidson, Donald/Harman, Gilbert (eds.), Reidel, Dordrecht/Boston (1972) 545–665
8. McCawley, James D.: Lexical Insertion in a Grammar without Deep Structure. Papers from the 4th Regional Meeting of the Chicago Linguistic Society (1968) 71–80
9. Moens, Marc, Steedman, Mark: Temporal Ontology and Temporal Reference. Computational Linguistics **14(2)** (1988) 15–28
10. Osswald, Rainer: On Result Nominalization in German. Proceedings of Sinn und Bedeutung 9 (2005) 256–270
11. Reckman, Hilke, Cremers, Crit: Deep parsing semantic interpretation of nominalizations and their expressed and unexpressed arguments. Leiden Working Papers in Linguistics **4(1)** (2007) 40–55
12. Scheffler, Tatjana: Nominalization in German. Unpublished Manuscript, University of Pennsylvania (2005)
13. Shin, Soo-Song: On the event structure of *-ung* nominals in German. Linguistics **39(2)** (2001) 297–319
14. Spranger, Kristina: Combining Deterministic Processing with Ambiguity-Awareness – The Case of Quantifying Noun Groups in German. PhD Thesis, University of Stuttgart (2006)
15. Vendler, Zeno: Facts and Events. In: Linguistics in Philosophy, Cornell University Press, Ithaca (1967) 122–146