

Knowledge-driven compound interpretation. A corpus study on German complex nouns headed by *-stoff*

Olav Mueller-Reichau, University of Leipzig, Leipzig, Germany
Matthias Irmer, OntoChem GmbH, Halle (Saale), Germany

This paper presents a case study on German compounds which have -stoff ('stuff /material/matter') as their head. The basic claim is that the head noun narrows down the range of interpretations by introducing a pragmatic scaffold which we pin down as a conceptual script. We obtained a data set by querying a large corpus selecting 4,026 nominal Stoff-compound lemmas. Using the 100 most frequent compounds as our development set, we observed 72 substance-denoting, five fabric-denoting and 11 knowledge-denoting compounds (corresponding to the basic meanings of Stoff in isolation according to standard dictionaries). From the observation that substance compounds (e.g. Farbstoff 'colouring agent'; Impfstoff 'vaccine') cluster around two events, synthesis and application, we induce a script consisting of these two event frames connecting three kinds of states. A Stoff-compound may thus be expected to denote an entity which is either the starting material of a synthesis (Rohstoff 'raw material'), an addendum in synthesis (Süßstoff 'sweetening agent'), the product of synthesis and, at the same time, an occasion for application (Kunststoff 'synthetic material'), or the instrument of application (Sprengstoff 'explosive'). As for fabric compounds (e.g. Seidenstoff 'silk fabric'), we argue that fabrication is a special kind of synthesis. Knowledge compounds (e.g. Lesestoff 'reading material') arguably result from a metaphoric transfer to the immaterial domain, with the two events acquisition (counterpart of synthesis) and contribution (counterpart of application). In this article, we test the obtained model against two test sets comprising 50 compounds with medium and low frequencies, respectively. Since it turns out that almost all test items successfully integrate into the script, our study provides support for pragmatic theories of compounding, according to which the interpretation of a compound is crucially guided by stereotype knowledge associated with the given compound's constituents.

Keywords: compound nouns, German, conceptual knowledge, frame semantics, word formation

1. Introduction

Nominal compounds fascinate as they successfully pair minimal syntactic structures with complex interpretations. With respect to syntactic structure building, (almost) all there is to say is that two units, *A* and *B*, undergo concatenation. This transparency on the formal side sharply contrasts with the situation on the side of meaning, where there are “virtually endless semantic opportunities” (Fleischer & Barz 2012: 130; our translation) of how *A* and *B* join with each other. This even leads to the often expressed opinion that the final interpretation of a compound is unpredictable (cf. Bauer 2017: 71). We believe and hope that such a radical conclusion must be wrong. If compound meanings were indeed unpredictable, the task of the linguist, which consists in modeling a mechanism that derives the interpretations that specific compounds have, would amount to doing the impossible. Such a model would have to predict the unpredictable. So we prefer to take it that the “relationship between the first and second elements [...] can receive just about any pragmatically conceivable interpretation” (Lieber 2016: 48). The question to consider is: What is a pragmatically conceivable interpretation.

To this end, this paper presents a case study on German compounds which have *-stoff* ('stuff /material/matter') as their *B*-constituent. What we would like to know is (i) whether the range of possible interpretations of an *A*-stoff compound is constrained, and (ii) if it is, then how. As just noted, we hold the view that (if we exclude abnormal contexts) question (i) should be affirmed. The idea that we pursue in this paper is that the head noun, *Stoff*, narrows down the range of possible interpretations by introducing a conceptual scaffold that, so to speak, channels the semantic contribution of the modifier. This channeling amounts to determining what is pragmatically conceivable.¹ The morphological structure of these compounds can be N+N (e.g. *Wasserstoff* 'hydrogen'), Adj+N (e.g. *Süßstoff* 'sweetening agent') and possibly V+N (e.g. *Klebstoff* 'adhesive').

The innovation of the present paper lies in the way we pin down the conceptual structure that we call "pragmatic scaffold" for a particular case, which we investigate item by item based on corpus data. We will offer a proposal, thereby answering the following research question:

- (1) **Research question:** What does the pragmatic scaffold that we postulate to underly compound meanings look like in the case of German Stoff-compounds?

Given these objectives, it seems natural to start by considering the meanings of the word *Stoff* when it is used as an autonomous noun. According to the DWDS, the German noun *Stoff* has three basic meanings. We present our own English translation here, see fn.2 for the original German text:²

- (2) **Basic meanings of Stoff:**

1. fabric made of natural or artificial fibers or yarns, which is sold in wide, very long webs and is processed especially into garments, bedding, curtains.
2. sth. that forms the matter for an artwork, for a scientific representation.

¹ That idea is, admittedly, not very innovative. The presence of some sort of cue which helps determining the interpretation of a complex word is particularly relevant for new words (Bauer 2017: 169). New words require that the relationship between the content of the modifier and the content of the head be traceable, i.e. transparently inferrable by language users. Otherwise the new word could not become accepted in the language community (Spencer 2011: 502).

² DWDS Digitales Wörterbuch der deutschen Sprache. Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart, hrsg. v. d. Berlin-Brandenburgischen Akademie der Wissenschaften, <https://www.dwds.de/>, accessed 2021-07-28:

Stoff, der

1. aus natürlichen oder künstlichen Fasern, Garnen hergestelltes Gewebe, das in breiten, sehr langen Bahnen in den Handel kommt und besonders zu Kleidungsstücken, Bettwäsche, Gardinen verarbeitet wird.
2. etw., das die thematische Grundlage für eine künstlerische Gestaltung, eine wissenschaftliche Darstellung bildet (Lehrstoff, Unterrichtsstoff, Gesprächsstoff).
3. etw. Gegenständliches, Körperhaftes, das sich durch räumliche Ausdehnung auszeichnet
 - a. (Wissenschaft, umgangssprachlich:) das Gegenständliche, Körperhafte in der objektiven Realität, Masse, Materie
 - b. (Wissenschaft:) einzelner, bestimmter, gegenständlicher, körperhafter Bestandteil der objektiven Realität
 - c. (salopp, verhüllend, übertragen:) alkoholisches Getränk
 - d. (salopp, verhüllend, übertragen:) Rauschgift

3. sth. objective, physical, which is characterized by spatial extension
 - a. (acad., colloq. :) the concrete, physical in the objective reality, mass, material
 - b. (acad. :) single particular physical part of objective reality
 - c. (sloppy, concealing, metaphoric :) alcoholic beverage
 - d. (sloppy, concealing, metaphoric :) drugs

Other resources deviate in some details, sometimes a subcategory is seen as a category unto itself, but overall there seems to be consensus about the tripartite classification.³ In this paper, we use the words “fabric”, “knowledge” and “substance” to reference the basic meaning categories 1 to 3, as they are represented in (2). Perhaps unsurprisingly, we will come across these three basic meanings again when we take a detailed look at nominal compounds whose head is formed by *-stoff*.

This paper is structured as follows: Section 2 lays down the theoretical background. Section 3 describes the used methodology. In Section 4, a model for the interpretation of *Stoff*-compounds is developed, which is then tested in Section 5. Concluding remarks will close the paper in Section 6.

2. Background

2.1 *The interpretation of a compound*

As a first approximation, we may summarize the sources of knowledge which feed the interpretation of a compound *AB* as follows:

- (3) **Ingredients of meaning in a compound**
 - a. There is the meaning of constituent A.
 - b. There is the meaning of constituent B.
 - c. Appearing together in a compound, the meanings of A and B become related to each other, whereby
 - d. the relation relevant in a given case is often dependent on context.

Of particular interest is, of course, the nature of the relation (3c). Many proposals of how to grasp that relationship have been made (see Bauer 2017 and Schäfer 2017 for overviews on the literature; Ingason & Sigurðsson 2020 for recent discussion). Following Bauer (2017: 72–73), we may sort them into two groups. Approaches of the first type work with a presupposed inventory of different basic semantic relations, out of which the compound may, so to speak, select one or the other relation, dependent on context. Let us call these approaches “semantic theories” here (in full awareness of the imperfection of this term). Approaches of the second type assume a single abstract semantic relation only, the precise content of which is determined by pragmatics. Such “pragmatic theories” (as we tentatively call them) assume that the

³ For other classifications see the following lexicographic sources:

<https://www.duden.de/rechtschreibung/Stoff>

<https://de.wiktionary.org/wiki/Stoff>

<https://de.thefreedictionary.com/Stoff>

<https://woerterbuchnetz.de/?sigle=DWB&lemma=Stoff>

relationship that links the meanings of *A* and *B* in a compound *AB* is not taken from independently given basic semantic relations, but derives from the specific meanings of *A* and *B* involved in the specific compound at hand.

The approach that we will defend in this paper is of the latter type. To explain the specific assumptions that we make, it is useful to give at least an impression of the kinds of explanations that have been offered in the literature. Fundamental work was done half a century ago, which is nevertheless still worth looking at directly. Thus, we will first sketch the theory of Levi (1978), which may count as a prototypical “semantic theory”. Then, we will introduce the theory of Fanselow (1981a, b), which would count as a “pragmatic theory” under our definition (but see below).

2.2 Levi 1978

Levi (1978) builds her analysis of compound nouns (complex nominals, CN, in her terminology) on a restricted set of very basic semantic relations. She calls them Recoverably Deletable Predicates (RDP) because they “may be deleted in the process of transforming an underlying relative clause construction in the typically ambiguous surface configuration of the CN” (Levi 1978: 76). Thus, for deriving an RDP, one starts from a paraphrase for the meaning of the compound noun which is then reduced to its basic predicate. For example, *tear gas* may be paraphrased by a relative clause construction as “gas that causes tears”, where the RDP is cause. Table 1 shows the full set of 9 RDPs together with what she calls the traditional terms for their meaning, as well as some of her examples (ibid: 76f).

Table 1: Levi’s Recoverably Deletable Predicates:

RDP	Traditional term	Examples
CAUSE	causative	tear gas
HAVE	possessive/dative	picture book
MAKE	productive; constitutive, compositional	honeybee; snowball use
USE	instrumental	steam iron
BE	essive/appositional	soldier ant
IN	locative (spatial or temporal)	field mouse; morning prayers
FOR	purposive/benefactive	horse doctor; arms budget
FROM	source/ablative	olive oil; test-tube baby
ABOUT	topic	tax law; abortion vote

Levi’s predicates seem to be rather general semantic predicates. For instance, the predicate IN subsumes both spatial and temporal as well as concrete and abstract location (Levi 1978: 83). The author defends her view against a more ne-grained taxonomy of relations, claiming that it is precisely her set of relations that determines the range of possible interpretations of most compound nouns.

Since its publication, Levi’s approach has received much attention and various extensions, but has also been criticized in various respects. See Schäfer (2017: 97–100) for a more detailed discussion. A major concern is the general problem that it is often possible to

analyze a compound noun in terms of a classification like RDPs in more than one way. A second criticism is that Levi's classification does not always capture well the full meaning aspect of compounds. As already Fanselow (1981b: 152f) points out, for instance, it is questionable whether USE is an appropriate relation for capturing the meaning of German *Polizeihund* ('police dog'), because it means more to be a proper *Polizeihund* than to merely be a dog used by the police.

Spencer (2011: 491) raises similar concerns with *elephant gun* being analyzed as a "gun for elephants." The correct level of granularity to be taken for choosing the set of meta predicates seems difficult to determine and ultimately arbitrary. Moreover, it seems desirable to go without these predicates altogether if we manage to derive the meaning of a compound from its parts by themselves, without resorting to independently stipulated predicates.

2.3 Fanselow 1981a,b

Above we characterized Fanselow's approach to German compounds as a pragmatic theory. In fact, however, according to the classification presented above, it is rather a hybrid of a semantic theory and a pragmatic theory. On the one hand, the author posits a set of primitive semantic relations that "have nothing to do with the meaning of the parts of the compound" (Fanselow 1981b: 138; our translation). The four relations are (a) "und" 'and', (b) "teil-von" 'is part of', (c) "lokalisiert" 'is localized at' and (d) "ähnelt" 'resembles'. According to Fanselow, they figure in the following compounds, for instance:

- (4)
- a. *Cola-Rum* 'rum and Coke'; *Hausboot* 'houseboat'
 - b. *Kugelschreiber* 'ball pen'; *Autokotflügel* 'car mudguard'
 - c. *Küstenstadt* 'coastal town'; *Bergdenkmal* 'mountain monument'
 - d. *Blutbuche* 'copper beech' (lit.: 'blood beech'); *Königstiger* 'bengal tiger' (lit. 'king tiger')

This is the "semantic" part of the theory. Now, on top of that, Fanselow assumes for all other compounds that the meaning relation does have to do with the meanings of the parts of the compound, i.e. that it can be derived from the idiosyncratic meanings of the constituents. Since these cases make the majority of German compounds (Fanselow 1981b: 156), it seems legitimate to call Fanselow's approach a "pragmatic theory".⁴

Both constituents *A* and *B* may serve as the source constituent, i.e. the one from which the relation is derived. Table 2 shows five examples for both possibilities, as analyzed by Fanselow (1981b: 156). We give the source constituent and the relation derived from it in bold face. This way, each line of the table may be read such that the bold-faced relation in the paraphrase stems from the meaning of the bold-faced part of the compound. For instance, on Fanselow's analysis, the relation "fire from" figuring in the meaning of *Raketenbasis* stems from (is a stereotype associated with) the meaning of *Rakete*, and the relation "grow in" figuring in the meaning of *Kornblume* derives from the meaning of *Blume*.

⁴ A reviewer disagrees with treating idiosyncratic meaning components of a word as pragmatic information. It seems that our view on conceptual knowledge deviates from his perspective. Unlike him (as it seems), we consider conceptual knowledge to be non-linguistic knowledge, although associated with a word. The issue clearly has fundamental philosophical implications, but as for the particular phenomenon investigated in this paper, nothing much hinges on this or that perspective on the topic.

Table 2: Fanselow's inferred relations

Example	Translation	Paraphrase/Relation
Raketenbasis	'missile base'	'base from which missiles are fired '
Taschenmesser	'pocketknife'	'knife to be carried in the pocket'
Fabrikgeige	lit. 'factory violin'	'violin produced in a factory'
Gartenblume	lit. 'garden flower'	'flower growing in the garden'
Teehaus	'tea house'	'house in which people drink tea'
Kornblume	'cornflower'	'flower growing in corn elds'
Gemüseladen	lit. 'vegetable shop'	'shop that sells vegetables'
Ölquelle	'oil spring'	'spring out of which comes oil'
Bücherregal	'bookshelf'	'shelf onto which books are put '
Automotor	'car engine'	'engine that drives a car'

The critical reader will immediately note some inconsistencies in Fanselow's conclusions. Some pieces of the relations expressed in the paraphrases do not seem to stem from the constituents which are identified as the source constituents. Consider, for example, the elements "from" and "in" in the paraphrases of *Raketenbasis* and *Teehaus*. This is considered a problem by the author himself (Fanselow 1981b: 202). It is, moreover, not clear why the relation "grow in" should stem from constituent A in *Gartenblume*, but from constituent B in *Kornblume*. Also, it appears to us that many (if not all) paraphrases which are said to derive from constituent A could be reconsidered in such a way that the relation will derive from constituent B. A pocketknife, for example, could plausibly be also described as a knife that is pocket-sized, i.e. that has a size that makes it fit in a pocket. This way put, we could argue that it is in fact the constituent *Messer* that supplies the relation ("has a size of") that is exploited for the interpretation of the compound, and that *Tasche* simply delivers a value for the size-attribute. Similarly, a base could be paraphrased as a location from which something is regularly sent off, with missiles being one such item that may be sent from a base. Like every artefact, furthermore, a violin is known to have been produced somewhere, but in the specific case of a violin it is also known that the place of production plays a crucial role for the quality of the instrument. It is thus not evident that the relation produced in the meaning of *Fabrikgeige* is supplied by *Fabrik*. It may also be supplied by *Geige*, with *Fabrik* merely specifying the place of production. As for *Teehaus*, we know about houses as such that life is going on inside of them. What precisely is going on inside may differ from house to house. So why not saying that *Tee* merely indicates what is going on inside a house called teahouse (drinking tea)? Then the relevant relation "go on inside of" would be delivered by the meaning of constituent B, *Haus*.

In the present paper, we present a case study on German compounds the *B*-constituent of which is *Stoff*. Upon consideration of the respective words that text corpora provide, we have been led to conclude that the interpretation of *Stoff*-compounds can successfully be described very much in the spirit of Fanselow (1981a,b), i.e. without appealing to independently given basic semantic relations. As we will show, the relation (3c) is always derivable from the

constituents of the respective *Stoff*-compound. It seems, moreover, that the relation is always derivable from the B-constituent, i.e. from the meaning of *Stoff*.

Note that we do not claim the general irrelevance of the basic semantic relations given in (4). The “und”-relation, for instance, is surely crucial to understanding coordinative compounds (Schäfer 2017: 113). However, we do not know a single coordinative *Stoff*-compound, and in this paper we are concerned with *Stoff*-compounds only.

2.4 *The proposal in general terms*

Generally speaking, we subscribe to the pragmatic view that the interpretation of compounds is guided by shared expectations about the role and function of the kinds of entities named in a given compound, as part of more general expectations about the course of the world. Expectations are geared towards a world proceeding along normal lines (cf. d’Avis 2016a,b on the linguistic relevance of normalcy expectations). These expectations, in turn, are defined by the conceptual background knowledge that speakers have and believe other speakers to have.

Given this, it is thus pieces of shared background knowledge that serve as the scaffold for speakers to determine the intended relationship between the content of the modifier (A) and the content of the head (*stoff*). Making structures of generalized background knowledge responsible for channeling interpretation is, of course, not a new idea. Similar proposals have been made, inter alia, by Hobbs et al. (1993) with respect to compounding, and by Plag, Andreou & Kawaletz (2018) with respect to derivation.

Fanselow (1981a,b) considers the relevant “pieces of shared background knowledge” to be stereotypes. For him, the relation that is not explicitly expressed in a compound is thus inferable from stereotypical knowledge associated with one of the two explicit constituents (the “source constituent”, as we called it above). Consider *Gemüseladen* ‘greengrocer’s shop’ from the examples given in Table 2. The head noun *Laden* ‘shop’ makes available the stereotype that shops are for selling things, and this relation “sells” is then exploited in the interpretation of the compound: if something is an instance of a *Gemüseladen*, it will be understood as a shop in which vegetables are being sold.

We basically agree with that view, but prefer to treat the “pieces of shared background knowledge” as conceptual frames, scripts or stories.⁵ Scripts can be conceived of as serialized event frames leading from one state to the next (Schank & Abel 1977; Busse 2012; Irmer & Mueller-Reichau 2018). Below we will argue that the use of the element *-stoff* as the head of a compound evokes a script involving two event frames connecting three kinds of states. Given this script, a *Stoff*-compound may be expected to denote an entity which is thematically involved either in the initial state or in the result state of one of these two events.

3. Methodology

Having laid out some theoretical background, we now come to present our study in detail. We obtained a data set by querying a large corpus and selecting all nominal compounds with a *B*-

⁵ This move enables us to connect compounding with central ideas of frame semantics according to which every use of a content word evokes a knowledge frame (Fillmore 1976, 1982; Fillmore & Baker 2009). Specifically, we follow the assumption that frames can be defined as recursive functional concepts (Barsalou 1992; Petersen 2007; Löbner 2015; Plag, Andreou & Kawaletz 2018).

constituent *Stoff*. This is described in 3.1. The method of semantic annotation is presented in Section 3.2.

In the sections to follow, a model for the interpretation of *Stoff*-compounds is first developed and then tested: In Section 4, we will use the 100 most frequent compounds as our development set for creating a model for their interpretation. In Section 5, the obtained model will be tested against two test sets comprising compounds with medium and low frequencies.

3.1 Data: corpus inquiry

Our starting point was to enquire the occurrences of *stoff in the public diachronic and synchronic German language resource DWDS. Specifically, we exploited a public reference corpus consisting of a broad variety of sources.⁶

We obtained our data sample by issuing the following query at the endpoint given in footnote 6:⁷

```
(5) COUNT (/^[A-Z](\w|-)*stoff[e|en|es|s]?/dg)
#BY [$1~s/(.*)[e|s]$/g] #DESC_COUNT
```

This query counts all tokens starting with a prefix consisting of an uppercase letter followed by any number of letters, numbers or hyphens (\w|-), and ending with ‘stoff’, possibly followed by inflection suffixes ([e|en|es|s]). Diacritic signs are included and only entire tokens are considered (/dg). These occurrences are then grouped by their lemmatized form and sorted descending by number of items. Lemmatization is done as provided by the corpus (\$) as far as it works, and additionally, a naive lemmatization by cutting off basic inflection suffixes ([e|s]) is performed. The number of remaining artefacts is quite low and can be neglected. No further refinement of the result is needed.

We obtained a result set of 4,026 lemmas, with a total of 121,710 occurrences. These show a typical Zipf distribution, with the most frequent lemma (*Rohstoff*) occurring 11,040 times (~9%), and 2,283 lemmas (~57%) with a single occurrence, so called “hapax legomena”. The full result set is given in the supplementary material (Mueller-Reichau and Irmer 2023).⁸

3.2 Semantic annotation

The semantic annotation of the data was performed in the following way. We considered each compound in isolation, that is to say, outside of any specific context (see Schäfer 2017: 229ff. for a similar practice). Both authors of this paper together decided on a paraphrase, guided by which interpretation was coming to mind first on an intuitive basis. In cases of doubt or disagreement, which for the 100 most frequent items were very few, we consulted standard resources, such as those mentioned in the introduction of this paper. We proceeded in two steps: First, we classified the compounds as belonging to one of the meaning classes substance, knowledge or fabric, recall (2). Then we stated for each compound a paraphrase consisting of

⁶ „Referenz- und Zeitungskorpus, Digitales Wörterbuch der deutschen Sprache“, <https://www.dwds.de/d/korpora/public>, accessed 2022-01-27; total corpus size: 1,33 billion tokens

⁷ Endpoint: <https://kaskade.dwds.de/dstar/public/dstar.per1>, accessed 2022-02-04; Query syntax: <https://kaskade.dwds.de/~moocow/software/ddc/querydoc.html>, accessed 2022-02-04

⁸ <https://doi.org/10.5281/zenodo.7577641>

the noun “substance”, “knowledge” or “fabric”, respectively, followed by a relative clause. We tried to state the relative clause such that it contains the A-constituent of the compound.

When the compound was unknown to both of us, we had to look up its meaning as well, of course. Especially for many hapax legomena, to which we turn below, it was necessary to consult the specific context in which they were mentioned.

The further we got into it, the more it became clear to us that a large fraction of the data (namely the substance-denoting compounds) had meanings that cluster around a small set of different kinds of events. These events showed up again and again in the meanings of the relative clauses. We concluded that these events provide the cornerstones of the scaffold we are after. Pinning down the nature of these events by choosing a meta predicate turned out to be difficult, however. At first we worked with three labels DISCOVERY, ADDITION and APPLICATION to cover the observed event structures. After reconsideration, we ended up with SYNTHESIS and APPLICATION, which we will introduce below in more detail.

For a few cases, our introspection-based methodology did not yield unequivocal results because two readings came to mind more or less equally well. In such cases, we provided two paraphrases distinguishing the relevant compounds by numbers (e.g. *Zellstoff 1* vs. *Zellstoff 2*). Note that we do not claim that these were the only cases of ambiguity. Needless to say, contextualization may trigger many more readings to come to mind than those that came to our minds at first glance.

4. Developing a model

In this section, we will develop a model to explain attestable interpretations of *Stoff*-compounds. Considering the 100 most frequent items, we will note that these compounds fall into three denotational categories. They may denote either substances, or knowledge, or fabrics. These classes obviously correspond to the three main classes established for the lexical noun *Stoff*, recall (2).

In Section 4.1, we isolate the by far largest group, substance-denoting *Stoff*-compounds, and we will give paraphrases for the (most salient) meanings that these items have in the null context. We will observe that substance-denoting compounds further subdivide into three categories: natural substances, substances involved in a synthesis, and substances relating to an application (to be explained below).

From the clusters observable in the paraphrases, we will draw the conclusion that a model of the scaffold guiding the interpretation of compounds headed by *Stoff* should contain two events, which are sequentially ordered to form a script. We call these events SYNTHESIS and APPLICATION, respectively, and we will offer a concrete proposal of how this script looks like, to be introduced in Section 4.2.

In Sections 4.3 and 4.4 we move on to ask how fabric and knowledge readings fit the script, which so far has been induced on the basis of substance readings alone. We will propose that fabric compounds are a species of substance compounds, and that knowledge compounds relate to our script by a metaphorical shift from the material domain to the immaterial domain.

4.1 *Substance readings*

We first looked at the 100 most frequent items on our result list. This decision might irritate because highly frequent items are prone to semantic shifts. We agree, but frequent compounds

are at the same time the ones that are most often experienced by speakers, thus probably exerting most influence on the shape of the pragmatic scaffold that we are after.

Recall that we consider the scaffold to be a frame. Frames are structures of prototypical knowledge (Busse 2012, 2018). As such, they are usage-based, i.e. sensitive to frequency. It therefore seems reasonable to appeal to the most frequent items in order to reconstruct the scaffold which we assume to channel compound interpretations. Our reasoning is that, if we can find the hypothesized conceptual scaffold behind the most frequent compounds the head of which is *Stoff*, and if that scaffold serves a model for the coinage of new words, as we suppose, then it should be possible to apply it to the compounds lower in frequency as well, as the latter are expected to show fewer idiosyncrasies due to lexicalization.

From the TOP 100 we excluded two obvious artefacts, i.e. the names *Christoff* and *Kristoff*. We also noticed 10 items that we may safely consider to be hyponyms of other items in the same list. These are: *Thioharnstoff*, *Sekundärrohstoff*, *Biotreibstoff*, *Luftsauerstoff*, *Luftschadstoff*, *Plastiksprengstoff*, *Blutfarbstoff*, *Biokraftstoff*, *Kernbrennstoff*, *Dieselmkraftstoff*. So we are left with 88 *Stoff*-compounds.

Within that sample, we find 17 compounds that name natural substances. A substance is a natural substance if it is simply present on planet Earth, existing independently of human manipulation. All natural substances in our list have been discovered by someone somewhere. The only exception is *Urstoff* ('primordial matter'), which has not been discovered, but is merely postulated to exist.⁹ If natural substances are named by *Stoff*-compounds, the A-constituent always points to some salient property that serves a mnemonic label for the respective class (Table 3).

In addition, there are 33 compounds that name substances that in one way or another participate in a synthesis. By this we mean, roughly, that different substances combine with each other to create a new substance. This includes, but is not limited to, chemical reactions. Often, the compound denotes some ingredient, the product or the base material of a synthesis event (Table 4).

Furthermore, we find 23 compounds that denote substances that are, so to speak, ready for utilization or application (Table 5). The definitional difference to compounds like *Süßstoff* ('sweetening agent') may be subtle, as sweetening may well also be counted as utilization. Nevertheless, there is the clear intuition that the substances listed in Table 4 participate in the creation of new substances, whereas those of Table 5 do not. Instead, the latter will, under reasonably foreseeable usage, be applied to cause a (usually intended) effect within the world of individuals.

4.2 *The script*

We saw that we can identify three different classes of substances denoted by *Stoff*-compounds within our data set. Natural substances are simply there, found in nature, and named by the respective compound expression. Other compounds indicate that the denoted substances play a role in processes of synthesization, which lead to the creation of new substances. The third class denoted by *Stoff*-compounds is made up of substances that serve a certain purpose independent of synthesization; they may be applied in order to reach a specific goal.

⁹This could be counted as an argument not to count *Urstoff* as the result of compounding.

Table 3: Compounds denoting natural substances

Compound	Paraphrase
Sauerstoff	substance available with property P (acid-forming)
Wasserstoff	substance available with property P (water)
Kohlenstoff	substance available with property P (carbon)
Stickstoff	substance available with property P (suffocating)
Kohlenwasserstoff	substance available with properties P (carbon) and Q (water)
Harnstoff	substance available with property P (found in urine)
Giftstoff 1	substance available with property P (poison)
Botenstoff	substance available with property P (deliver message in organism)
Eiweißstoff	substance available with property P (protein)
Schwefelwasserstoff	substance available with properties P (sulfur) and Q (water)
Urstoff	substance from which all natural substances have developed
Pflanzenstoff	substance available with property P (in plants)
Lebensstoff	substance available with property P (enabling life)
Schwefelkohlenstoff	substance available with properties P (sulfur) and Q (carbon)
Fluorchlorkohlenwasserstoff	substance available with properties P (fluorine), Q (chlorine), R (carbon) and S (water)
Krankheitsstoff	substance available with property P (disease causing)
Signalstoff	substance available with property P (transmit signal in organism)
Mineralstoff	substance available with property P (have minerals)

The three classes of substances relate to each other in a natural way, which is neatly describable as a script:

[A] script is a predetermined, stereotyped sequence of actions that denote a well-known situation. [...] Each action results in conditions that enable the next to occur. To perform the next act in the sequence, the previous acts must be completed satisfactorily. (Schank & Abelson 1977: 45, 1975: 152).

In accordance with that definition, we may conceive of the three classes of substances as being involved in a “stereotyped sequence of actions”: Natural substances, once discovered, are simply there with certain distinguishing properties, available to be subjected to a synthesis event or “action” thereby bringing about new substances with new properties, which, unless further synthesized, may in turn be used in an application event or “action”. This, we claim, is the scaffold that guides expectations about possible interpretations.

Figure 1 thus shows the script that is arguably evoked by the use of a German compound which is headed by *Stoff*. It consists of three states (symbolized by circles), which are connected

with each other by two events (symbolized by boxes). Each state may be the occasion for an event (indicated by the label “enables”), or it may be the result of an event (indicated by “result”). For more details on the structure and formal definition of scripts, we refer the reader to Irmer & Mueller-Reichau (2018), due to space constraints.

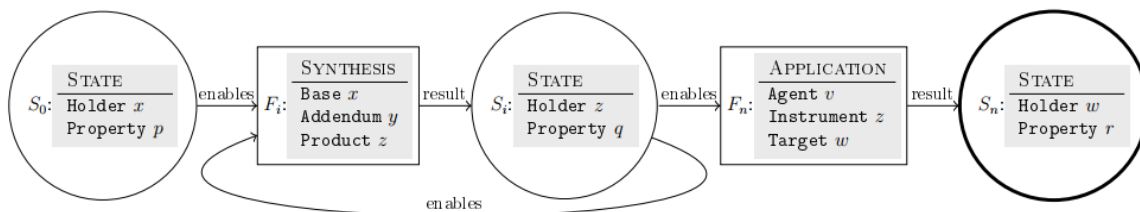


Figure 1: The Stoff-script

Any *Stoff*-compound will, according to our proposal, determine its denotation relative to the script in Figure 1. Below we will argue this holds not only for substance-denoting compounds, but also for fabric-denoting and knowledge-denoting compounds. To determine denotation relative to the script means that the compound is a naming strategy for a substance (or fabric, or piece of knowledge) that exploits the fact that speakers share stereotypical knowledge about what is denoted by *Stoff*. The modifier constituent *A* of the compound thereby indicates the relevant aspect of the shared *Stoff*-script.

Let us look at some examples in detail. We will start with *Impfstoff* (‘vaccine’). This compound has been paraphrased as naming a “substance that is an instrument in APPLICATION causing immunity” (cf. Table 5). The A-constituent *Impf-* points to an event of VACCINATION, which is a subkind of the more general APPLICATION. In our script of Figure 1, *Impfstoff* denotes an entity *z* playing the role of Instrument in an APPLICATION event. The specification of APPLICATION as VACCINATION contributes the information that the Holder *w* of its result state *S_n*, which is the Target of the APPLICATION event (i.e. the person being vaccinated), has the Property of being immune. Although other parts of the script are not explicitly addressed, they will be implicitly understood: A state *S_i* enabling such an APPLICATION is at the same time a state resulting from a SYNTHESIS which has *z* as its Product. As an example for the subset of *Stoff*-compounds naming natural substances (cf. Table 3), *Sauerstoff* (‘oxygen’) names a substance which is available with an attributed property *p* of being acid-forming. Note that the essential property for the evolution of acids had been erroneously attributed to oxygen at the time of coining of the German term in the 18th century; later it was found out that it is in fact hydrogen, rather than oxygen, which is responsible for acidity.¹⁰ In our script, *Sauerstoff* denotes an entity *x* that is Holder of a state *S₀* having (or being attributed) the Property *p* of acid-forming. This state enables a SYNTHESIS event to occur, where *x* fills the role of its Base (or Addendum, depending on the perspective taken).

¹⁰ <https://de.wikipedia.org/wiki/Sauerstoff#Geschichte> or <https://en.wikipedia.org/wiki/Oxygen#Etymology>

Table 4: Substance-denoting compounds relating to SYNTHESIS

Compound	Paraphrase
Rohstoff	substance that is raw in the sense that it has not yet undergone SYNTHESIS
Ausgangsstoff	substance that is base (point of departure) of a SYNTHESIS
Naturstoff	substance that is natural in the sense that it has not yet undergone SYNTHESIS
Grundstoff	substance that is base of SYNTHESIS
Farbstoff	substance that is ingredient in SYNTHESIS adding colour
Süßstoff	substance that is ingredient in SYNTHESIS adding sweetness
Duftstoff	substance that is ingredient in SYNTHESIS adding smell
Inhaltsstoff	substance that is ingredient in SYNTHESIS
Zusatzstoff	substance that is an additional ingredient in SYNTHESIS
Zellstoff 1	substance that is product of SYNTHESIS available with property P (cell)
Schaumstoff 1	substance that is product of SYNTHESIS available with property P (foam)
Kunststoff	substance that is product of SYNTHESIS available with property P (artificial)
Faserstoff	substance that is product of a SYNTHESIS available with property P (fibre)
Wertstoff	substance that makes a valuable contribution to SYNTHESIS
Ersatzstoff	substance that is a substitute ingredient in SYNTHESIS
Gerbstoff	substance that is ingredient in SYNTHESIS tanning skin into leather
Aromastoff	substance that is ingredient in SYNTHESIS adding flavor
Konservierungsstoff	substance that is ingredient in SYNTHESIS adding extended longevity
Glanzstoff	substance that is ingredient in SYNTHESIS adding reflective character
Abfallstoff	substance that is a useless by-product of SYNTHESIS
Fremdstoff	substance that is a misplaced ingredient of SYNTHESIS
Füllestoff	substance that fills up the amount of an ingredient of SYNTHESIS
Arzneistoff	substance that is ingredient of SYNTHESIS making the product medicine
Bitterstoff	substance that is ingredient in SYNTHESIS evoking bitter taste
Spaltstoff	substance that is ingredient in SYNTHESIS (responsible for fission in a nuclear reaction)
Hemmstoff	substance that is ingredient in SYNTHESIS (prevents or decreases the rate of given reaction)
Lichtstoff	substance that is ingredient in SYNTHESIS causing light
Käsestoff	substance that is the milk ingredient in SYNTHESIS (cheese production) that turns into cheese
Zeugungsstoff	substance that is ingredient in SYNTHESIS (procreation)
Werkstoff	substance that is ingredient in SYNTHESIS (crafting sth)
Baustoff	substance that is ingredient in SYNTHESIS (building sth)
Hilfestoff	substance that is ingredient in SYNTHESIS ensuring cohesion of product
Explosivstoff	substance that can produce SYNTHESIS (explosion) when released suddenly

Table 5: Substance-denoting compounds relating to APPLICATION

Compound	Paraphrase
Sprengstoff	substance that is instrument in APPLICATION (blowing something up)
Treibstoff	substance that is instrument in APPLICATION (driving an engine)
Wirkstoff	substance that is instrument in APPLICATION causing an intended effect
Brennstoff	substance that is instrument in APPLICATION obtaining energy by burning
Impfstoff	substance that is instrument in APPLICATION (vaccination) causing immunity
Kraftstoff	substance that is instrument in APPLICATION generating power to drive an engine
Schadstoff	substance that is instrument in APPLICATION doing damage
Zündstoff	substance that is instrument in APPLICATION igniting an engine
Nährstoff	substance that is instrument in APPLICATION to feed an organism with nutrients
Kampfstoff	substance that is instrument in APPLICATION fighting a military enemy
Klebstoff	substance that is instrument in APPLICATION (gluing something)
Giftstoff 2	substance that is instrument in APPLICATION causing intoxication
Zellstoff 2	substance that is instrument in APPLICATION causing absorption
Wärmestoff	substance that is instrument in APPLICATION evolving heat
Schaumstoff 2	substance that is instrument in APPLICATION minimize force of impact
Nahrungsstoff	substance that is instrument in APPLICATION to feed an organism with nutrients
Dämmstoff	substance that is instrument in APPLICATION (supplying insulation)
Schmierstoff	substance that is instrument in APPLICATION (lubricating an engine)
Ballaststoff	substance that is instrument for APPLICATION (provide organism with fibres)
Betriebsstoff	substance that is instrument in APPLICATION (driving a machine)
Abwehrstoff	substance that is instrument for APPLICATION causing protection of organism
Reststoff	substance that remains after APPLICATION
Lockstoff	substance that is instrument in APPLICATION (attract someone or something)
Reizstoff	substance that is instrument in APPLICATION (harm by irritating organs)

Regarding substance-denoting *Stoff*-compounds related to SYNTHESIS, we take *Duftstoff* (‘scent’, ‘aromatic substance’) as an example. This compound is paraphrased as referring to a substance that is an ingredient in SYNTHESIS adding smell (cf. Table 4). In terms of Figure 1, *Duftstoff* denotes an entity y which is an Addendum of a SYNTHESIS event. In the result state S_i of that event, the resulting Product z has the Property of having a certain smell.

This way, the range of possible interpretations is narrowed down considerably, but of course there is also still quite a bit of space for interpretive variability. Dependent on context, one and the same compound expression may gain different interpretations. Our script-based approach makes specific predictions in this regard. We may, for instance, expect to find

compounds that, if interpreted out of the blue, give rise to an ambiguity between being the product of a synthesis or the occasion of an application. Consider, for instance, *Zellstoff* ‘chemical pulp’ (see above).¹¹ We may also find three-way ambiguities. Take *Giftstoff* ‘toxin’, which can either refer to a natural substance with the property of being toxic (6a), to the product of a synthesis which releases toxic side-products (6b), or to a toxic substance which is used for a specific application (6c).

- (6) a. *Mykotoxine (Schimmelpilzgifte) und Pflanzentoxine (z. B. Pyrrolizidinalkaloide) sind natürliche Giftstoffe, die Lebensmittel kontaminieren können. (“Mycotoxins and plant toxins are natural toxins that can contaminate food.”)*¹²
- b. *Der Nachteil von Vinyl ist, dass es nicht natürlich und ziemlich umweltschädlich ist. Bei seiner Herstellung entstehen Giftstoffe. (“The disadvantage of vinyl is that it is not natural and quite harmful to the environment. During its production, toxins arise.”)*¹³
- c. *Auch nach Stalin setzte der KGB Giftstoffe gegen politische Gegner ein. (“Even after Stalin the KGB continued to apply toxins against political opponents.”)*¹⁴

Why is it an advantage to distinguish between these three readings, instead of simply paraphrasing *Giftstoff* as “substance that has the quality of being toxic” and relegating all the specific information to cues found in the local context, i.e. *kontaminieren* in (6a), *entstehen* in (6b) and *setzte ... ein* in (6c), as suggested by reviewer? Well, while admittedly also the verb contributes its part of information (see Irmer & Mueller-Reichau 2018 for a sketch how frame blending might look like), the contribution of the noun can be narrowed down to precisely one of these three readings, nothing more or less. Our account constrains the context for interpreting compounds considerably by providing specific slots in a script to be filled with the entity denoted by such a compound.

4.3 Fabric readings

We have derived the script in Figure 1 on the basis of substance-denoting compounds only. These sum up to 72 items in our sample given in the Tables 3 to 5 (note that three compounds were coded by two paraphrases each, i.e. *Giftstoff*, *Schaumstoff* and *Zellstoff*). To these, we should add the ten hyponyms, and the two artefacts *Kristoff* and *Christoff*. This way we arrive at 84. What about the remaining 16?

¹¹ The fact that the paraphrase of *Zellstoff* 2 does not contain the word cell might indicate that the compound once has extended its original synthesis-related meaning to develop an application-related reading.

¹² <https://www.lebensmittelverband.de/de/lebensmittel/sicherheit/unerwuenschte-stoffe-kontaminanten>

¹³ <https://www.homelife-guide.com/childrens-bedroom-floor-options-1314791>

¹⁴ <https://www.deutschlandfunk.de/morden-nach-system-der-fall-nawalny-und-die-tradition-100.html>

Table 6: Fabric-denoting compounds

Compound	Paraphrase
Seidenstoff	fabric that is product of SYNTHESIS (fabrication) with base material silk
Wollstoff	fabric that is product of SYNTHESIS (fabrication) with base material wool
Baumwollstoff	fabric that is product of SYNTHESIS (fabrication) with base material cotton
Futterstoff	fabric that is ingredient in SYNTHESIS (tailoring) serving as lining
Kleiderstoff	fabric that is ingredient in SYNTHESIS (tailoring a dress)

Five compounds display fabric readings; they denote fabrics (Table 6). These compounds will integrate into the picture we arrived at so far, if we consider the word “fabric” to be shorthand for the “product of a fabrication event” (which does not seem too far-fetched), and if we assume that fabrication is a special case of synthesization. Under this assumption the road is free to treat the denotations of *Seidenstoff*, *Wollstoff* and *Baumwollstoff* in line with Figure 1 as holders of a state which results from a synthesis (fabrication). Due to their location within the script, they come close in meaning to the cases of *Schaumstoff*, *Kunststoff*, *Zellstoff*, and *Faserstoff* in Table 4. The denotations of *Futterstoff* and *Kleiderstoff* can be viewed as ingredients in a synthesis/tailoring. We propose, in other words, that fabric-denoting compounds are ultimately substance-denoting, and that their meaning can accordingly be expressed in terms of the script given in Figure 1. *Seidenstoff* (‘silk fabric’), for instance, denotes an entity z that is Product of a SYNTHESIS which has silk as its Base material x . The result state S_n of this event is characterized by the availability of z having the Property q . In this case, q is the property of being made from silk.¹⁵

A different case of a fabric-denoting compound is represented by *Kleiderstoff* (‘dress fabric’). This compound denotes an entity z that is Ingredient of a SYNTHESIS event. Due to the presence of the A -constituent *Kleider-*, the event is more specifically understood as an instance of TAILORING a dress. In the result state of this event, its Target w has the Property r of being a dress.

4.4 Knowledge readings

Among the TOP 100 *Stoff*-compounds, there are still 11 knowledge-denoting items. Note that one of them, *Filmstoff*, is duplicated in Table 7, because out of the blue both paraphrases seem to equally qualify as interpretation of the compound. With respect to the compounds in Table 7, we propose that they come about from a metaphorical transposition from the material domain of substances into the immaterial domain of information/knowledge.¹⁶ The processing of material substances (synthesization) that we analyze as being relevant for substance-denoting *Stoff*-compounds finds its immaterial counterpart in the processing of information, i.e. in someone's acquisition of knowledge. And similar to compounds that denote substances usable

¹⁵ A reviewer suggested that “fabric made of silk” would be a simpler paraphrase for grasping the meaning of *Seidenstoff* and that our approach would add unnecessary complexity. However, we think that by grounding the meaning in the script we gain an explanation for where the relation dubbed as “made of” actually comes from, and that we do not complicate the picture, but elaborate it by explicating the event which is implicit in the relation “made of”.

¹⁶ Thanks to Matthias Hüning for encouraging us to defend that claim.

to impact on the physical world (application), we find a group of knowledge-denoting compounds usable to impact on communication, i.e. to make a contribution to ongoing discourses. In the respective paraphrases we accordingly use ACQUISITION and CONTRIBUTION as metapredicates. We thus use “synthesis” for the processing of substances and “acquisition” for the processing of information, and we use “application” for making a material intervention (in the physical world of substances) and “contribution” for making a verbal intervention (in the immaterial sphere of knowledge states).

For the sake of explicitness, let us take *Lesestoff* (‘reading material’) as an example. It is paraphrased as ‘knowledge that is content in ACQUISITION by reading.’ A script in the domain of information and knowledge could be shaped similar to the one of Table 1 basically as a sequence of an ACQUISITION and a CONTRIBUTION event. Given this, the compound *Lesestoff* is analyzable as denoting an entity x that is the Content of a general ACQUISITION event which is narrowed down to READING by the A-constituent *Lese-*. Once acquired through reading, the information turns into knowledge and is available as a Theme for a subsequent CONTRIBUTION.

As a second example, *Diskussionsstoff* (‘material for discussion’) is a compound related to the second event in the script. It has been paraphrased as “knowledge that is the theme of a CONTRIBUTION triggering a discussion.” In the knowledge-related *Stoff*-script, it accordingly plays the role of the Theme in a CONTRIBUTION event which has a discussion as its result.

Table 7: Knowledge-denoting compounds

Compound	Paraphrase
Unterrichtsstoff	knowledge that is content in ACQUISITION during a lesson
Lesestoff	knowledge that is content in ACQUISITION by reading
Lehrstoff	knowledge that is content in ACQUISITION from teacher
Lernstoff	knowledge that is content in ACQUISITION by learning
Filmstoff 1	knowledge that is content in ACQUISITION by watching a movie
Wissensstoff	knowledge that is content in ACQUISITION
Schulstoff	knowledge that is content in ACQUISITION in school
Bildungsstoff	knowledge that is content in ACQUISITION to enhance education
Gesprächsstoff	knowledge that is the theme of CONTRIBUTION triggering conversation
Konfliktstoff	knowledge that is the theme of CONTRIBUTION triggering a conflict
Diskussionsstoff	knowledge that is the theme of CONTRIBUTION triggering a discussion
Filmstoff 2	knowledge that is available as theme of a movie

In the knowledge-related *Stoff*-script, it accordingly plays the role of the Theme in a CONTRIBUTION event which has a discussion as its result.

The case of *Filmstoff 2* can be seen as referring to information that is “simply there”, available as the subject of some medium (here: of a movie), ready to be acquired as knowledge.

This, in a way, parallels the class of natural substances. Similar examples below the Top 100 are *Romanstoff* ('theme of a novel') or *Märchenstoff* ('theme of a fairy tale').

5. Testing the model

In this section, we will test whether the model developed so far will prove successful in accounting for the meanings of *Stoff*-compounds lower in frequency. The first test set will be 50 compounds from the lower end, i.e. 50 hapax legomena randomly chosen from our corpus (Section 5.1). The second test set will be chosen from the middle of our corpus, i.e. 50 compounds with ranks 501 to 550 in frequency (Section 5.2). Results will be presented and discussed in Section 5.3.

5.1 Test set LAST50

To test our hypothesis that the range of possible interpretations of *Stoff*-compounds is constrained by the script structure in Figure 1, we will first look at the lower end of the data set obtained. Recall from above that more than half (2,283) of all 4,026 of *Stoff*-compound lemmas are hapax legomena. Assuming that compounds that are new to conversants must have either a transparent or otherwise retrievable interpretation at the time of creation, we may expect their meanings to fit within the proposed model equally well or even better than high-frequency items, which may already show idiosyncracies due to lexicalization.

To check whether this expectation is borne out, we randomly selected 50 items out of the 2,283 hapax legomena in our data set. Of these, 26 items have more frequently attested hyperonyms which we dealt with already elsewhere, which is why they do not appear in Table 8 and Table 9. One item, *Kästner-Kinderbuchstoff*, has a hyperonym within the set of hapax legomena. The respective hyponyms are *Beruhigungswirkstoff*, *Luftschadstoff*, *Spezialwirkstoff*, *Mohn-Wirkstoff*, *Biodieselmotorenkraftstoff*, *Rindeninhaltsstoff*, *Holzinhaltstoff*, *Pasteur-Impfstoff*, *Atom-Rohstoff*, *Schmerzmittelwirkstoff*, *Schönheitswirkstoff*, *Hydrazonfarbstoff*, *Hauptlernstoff*, *Pestizid-Grundstoff*, *Koalitionssprengstoff*, *Flüssig-Kraftstoff*, *Fettgewebes-Botenstoff*, *Bewusstseins-Treibstoff*, *Düsenwerkstoff*, *Fluoreszenzfarbstoff*, *Papierfarbstoff*, *Stilbenfarbstoff*, *Gummi-Rohstoff*, *Arzneimittelgrundstoff*, *Kästner-Kinderbuchstoff*, *Filmlockstoff*, and *Holzspan-Werkstoff*. There is just one artefact, namely *Molostoff*, which is a germanized version of the Slavic name of General Molostov.

We determined the senses of the remaining items by performing Google queries, thereby consulting various resources, ranging from entries in encyclopaediae and scientific publications to online offers for sale and private websites. Results are shown in Table 8.

With these senses as target meanings, we then stated paraphrases following the procedure described above. It turned out that almost all (see below) items have interpretations that easily meet the script, as can be seen in Table 9: we found seven substance compounds, 10 fabric compounds and six compounds having an immaterial meaning. Two items, *Türkenstoff* and *Hightechstoff*, were categorized differently by us because we could establish two readings for each of them.

There are only two problematic cases, namely *Erscheinungstoff* and *Gedächtnisspeicherstoff*. These are the only data that we think resist a smooth integration into our system.

5.2 Test set MID50

As a second test set, we extracted 50 compounds from the medium frequency range, those with ranks 501 to 550 according to lemma frequency. A major part (21 compounds) are hyponyms of TOP 100 compounds, they thus have already been shown to integrate into the proposed script. These are *Biobrennstoff*, *Blutersatzstoff*, *Chemiekampfstoff*, *DNA-Impfstoff*, *Fckw-Ersatzstoff*, *Festkraftstoff*, *Fettersatzstoff*, *Flugkraftstoff*, *Flüssigstickstoff*, *Gesamtkohlenstoff*, *Grundrohstoff*, *Importrohstoff*, *Influenza-Impfstoff*, *Monomethylolharnstoff*, *Nitrofarbstoff*, *Nuklearsprengstoff*, *Radiokohlenstoff*, *Sechsfachimpfstoff*, *Ökokraftstoff*, *Alizarinfarbstoff*, and *Baurestoff*.

The remaining items are represented in Table 10 and Table 11. The first table gives approximate translations of the German words, the second shows our paraphrases. There was no difficulty in paraphrasing their meanings according to the developed model.

5.3 Results and Discussion

Nearly all items of the MID50 and LAST50 sets could be categorized as belonging to a certain phase of the script developed from the items in the TOP100 set. The distribution of readings and script phases in the three corpus sample sets is shown in Table 12.

In the TOP100, we have a vast majority of substance readings (83%), very few fabric readings (5%) and some more knowledge readings (12%). Regarding the location in the script, 21% refer to materials in the initial state of the script (referred to as AVAILABILITY in the table). Within substance and fabric readings, 39% fill a slot in a SYNTHESIS frame, and 29% play a role in an APPLICATION. As for knowledge readings, we observe 8% playing a role in ACQUISITION and 3% evoking CONTRIBUTION.

Similar patterns can be observed in the MID50 data. Regarding the distribution of readings, we have a slightly decreased percentage of substance readings (70%), a bigger proportion of fabric readings (22%), and somewhat less knowledge readings (10%). At least for substance and fabric readings, in MID50 the distribution into the different positions in the script is very similar to TOP100 (22% AVAILABILITY, 44% SYNTHESIS, and 28% Application). For knowledge readings, there is an equal distribution between ACQUISITION and CONTRIBUTION (both 4%).

The LAST50 data set shows a decreased share of substance readings (67%) and more fabric and knowledge readings (20% and 16%, respectively). Script positions are less frequent at the beginning of the script (16%), but more frequent in relation to SYNTHESIS (49%) and in relation to APPLICATION (33%). The shares of ACQUISITION (4%) and CONTRIBUTION (2%) are similar to the other data sets.

As a general observation on the distribution of readings, one can see a monotonically decreasing frequency of substance readings from TOP100 over MID50 to LAST50, while the shares of fabric readings, and somewhat less clearly also knowledge readings, increase. Regarding script phase, the proportion of script-initial positions (AVAILABILITY) is higher in MID50 and lower in LAST50. The share of SYNTHESIS-related meanings gets bigger with falling frequency (from 39% over 44% to 49%), while APPLICATION is slightly more frequent in LAST50. Thus, the script phase seems to be geared towards the middle of the script with decreasing frequency of *Stoff*-compounds. Unsurprisingly, the share of compounds which are hyponyms of other compounds increases from TOP100 (22%) over MID50 (42%) to LAST50 (56%). Over all three sets we observe a similar ratio between the shares of

AVAILABILITY, SYNTHESIS and APPLICATION, and, even if the data for knowledge readings is sparse, also between ACQUISITION and CONTRIBUTION. The complete data sets and results are given in the supplementary material (see footnote **Chyba! Záložka nie je definovaná.** above).

Table 8: Hapax legomena, senses

Compound	Sense
Erscheinungsstoff	the bearer of sth. immaterial showing up in the material world
Baiaststoff	a typo or variant of the frequent term <i>Ballaststoff</i> ('fibre')
Blaudruckstoff	blue print fabric
Schädlichkeitsstoff	substances that are harmful for the organism (= <i>schädlicher Stoff</i>)
Kennzeichnungsstoff	marking agent, labeling substance (= <i>Markierstoff</i>)
Botanikstoff	fabric having a floral design
Halbmondstoff	fabric having a night sky design
Brokatstoff	fabric that has has ornamental brocades
Mittelalterstoff	fabric that is like fabric was in medieval times
Struckstoff	upholstery fabric, a low quality substitute for leather
Hightechstoff	high technology manufactured material
Polierstoff	fabric to be used for polishing
Inlettstoff	fabric for a ticking (the tightly woven cover of a pillow)
Bettbezugsstoff	fabric out of which a bed cover is made
Kinderbuchstoff	a story which is suitable for a children's book
Revuestoff	a story which is suited to become the topic of a revue
Türkenstoff 1	a kind of heroine
Türkenstoff 2	the subject of a play which is thematically related to the orient
Gedächtnisspeicherstoff	memory storage substance in a cell
Lehrgangsstoff	knowledge that will be acquired by taking a training course
Ideenstoff	knowledge that serves the inspiration for creating something new
Jahrhundertstoff	a topic so important that people talk about it for a century
Molostoff	artefact: germanized version of Slavic name of General Molostov

Table 9: Hapax legomena, paraphrases

Compound	Paraphrase
Erscheinungsstoff ?	substance materializing something immaterial
Gedächtnisspeicherstoff ?	substance in a cell memorizing information
Baiastoff	substance that is ingredient in SYNTHESIS (digestion) without being broken down by enzymes
Kennzeichnungsstoff	substance that is ingredient in SYNTHESIS marking the product
Hightechstoff 1	substance that is the product of SYNTHESIS using high-tech technology
Türkenstoff 1	substance that is instrument in APPLICATION (drug consumption) sold by Turks
Schädlichkeitsstoff	substance that participates in APPLICATION doing harm to organism
Struckstoff	fabric that is the ingredient in SYNTHESIS (manufacturing furniture) to upholster
Inlettstoff	fabric that is ingredient in SYNTHESIS (tailoring a ticking)
Bettbezugsstoff	fabric that is ingredient in SYNTHESIS (tailoring a bed cover)
Blaudruckstoff	blue fabric that is the product of SYNTHESIS (fabrication) involving a special colouring technique
Botanikstoff	fabric that is the product of SYNTHESIS (fabrication) having a floral design
Halbmondstoff	fabric that is the product of SYNTHESIS (fabrication) having a night sky design
Brokatstoff	fabric that is the product of SYNTHESIS (fabrication) having ornamental brocades
Mittelalterstoff	fabric that is the product of SYNTHESIS (fabrication) using medieval technology
Hightechstoff 2	fabric that is the product of SYNTHESIS (fabrication) using high-tech technology
Polierstoff	fabric that is instrument in APPLICATION (polishing sth.)
Kinderbuchstoff	knowledge that is available as the subject of a children's book
Revuestoff	knowledge that is available as the subject of a revue
Türkenstoff 2	knowledge that is available as the subject of a play about Turks
Ideenstoff	knowledge that is available as the initiation of a new idea
Lehrgangsstoff	knowledge that is content in ACQUISITION by taking a training course
Jahrhundertstoff	knowledge that makes a CONTRIBUTION triggering century-long conversations

Table 10: Compounds of medium frequency, translations

Compound	Translation
Agrarstoff	agricultural material
Erdenstoff	“earth substance”: some component of soil
Pektinstoff	pectin
Samenstoff	sperm
Markstoff	myelin
Verbrauchsstoff	consumable good
Ganzstoff	“whole cloth”: name of the pulp that figures in paper production
Vorläuferstoff	precursor chemical, drug precursor
Würzstoff	substance used for seasoning
Ausscheidungsstoff	excretory substance
Bukettstoff	bouquet substances, characteristic odorants of the grape (winemaking)
Reinstoff	pure substance, composed of only one chemical compound or one chemical element
Sekundärstoff	phytochemical
Textilstoff	fabric, cloth
Zauberstoff	magic cloth; spellcloth (World of Warcraft)
Ermüdungsstoff	fatigue substance, substances that makes the muscles tired
Krawattenstoff	tie fabric
Tarnstoff	camouflage fabric
Verhüllungsstoff	fabric to cover sth
Anzugsstoff	suit fabric
Schirmstoff	umbrella fabric
Streifenstoff	fabric with stripes
Tweedstoff	Tweed
Glitzerstoff	fabric which has sparkles on it
Heldenstoff	heroic tales
Memorierstoff	memorization material, information to be memorized
Studiumsstoff	lecture materials, course materials
Verhandlungsstoff	negotiation material
Propagandastoff	propaganda material

Table 11: Compounds of medium frequency, paraphrases

Compound	Paraphrase
Agrarstoff	substance available with property P (relevant for agriculture)
Erdenstoff	substance available with property P (component of soil)
Pektinstoff	substance available with property P (like clot)
Samenstoff	substance available with property P (contains spermatozoa)
Markstoff	substance available with property P (forms the myelin sheath (= Markscheide))
Sekundärstoff	substance available as non-essential nutritive plant material
Verbrauchsstoff	substance that is necessary ingredient in SYNTHESIS but forms no part of the product
Ganzstoff	substance that is the interstage product of SYNTHESIS (paper manufacturing)
Vorläuferstoff	substance that is the base of a drug-producing SYNTHESIS
Würzstoff	substance that is an ingredient in SYNTHESIS (cooking) adding seasoning
Ausscheidungsstoff	substance that is the excretory product of SYNTHESIS (metabolism)
Bukettstoff	substance that develops during SYNTHESIS (winemaking) adding odorants
Reinstoff	substance that is a pure/homogeneous ingredient in SYNTHESIS
Ermüdungsstoff	substance that makes muscles tired in APPLICATION
Krawattenstoff	fabric that is the ingredient in SYNTHESIS (tailoring ties)
Tarnstoff 1	fabric that is the product of SYNTHESIS (fabrication) with camouflage pattern
Textilstoff	fabric that is the product of SYNTHESIS (fabrication)
Anzugsstoff	fabric that is the ingredient in SYNTHESIS (tailoring suits)
Schirmstoff	fabric that is the ingredient in SYNTHESIS (manufacturing umbrellas)
Streifenstoff	fabric that is the product of SYNTHESIS (fabrication) with stripes
Tweedstoff	fabric that is the product of SYNTHESIS (fabrication) having tweed texture
Glitzerstoff	fabric that is the product of SYNTHESIS (fabrication) having sparkles
Verhüllungsstoff	fabric that is instrument in APPLICATION (to cover sth.)
Zauberstoff	fabric that is instrument in APPLICATION (performing magic)
Tarnstoff 2	fabric that is instrument in APPLICATION (to camouflage)
Heldenstoff	knowledge that is the subject of a heroic tale
Memorierstoff	knowledge that is content in ACQUISITION via memorization
Studiumsstoff	knowledge that is content in ACQUISITION by learning lectures in academic studies
Verhandlungsstoff	knowledge that makes a CONTRIBUTION in negotiations
Propagandastoff	knowledge that makes a CONTRIBUTION for propaganda purposes

6. Conclusion

In this paper, we have provided support for Fanselow's (1981a,b) intuition that the interpretation of a German compound is crucially guided by stereotype knowledge associated with the compound's constituents. We contributed to advancing this perspective by undertaking a case study on compounds headed by the nominal element *Stoff*. By conducting an item-by-item analysis of attested *Stoff*-compounds we induced the knowledge structure underlying their interpretation, which we pinned down in terms of a script involving two serialized events.

Table 12: Distribution of readings and script phases in data sets

Reading	Script Phase	Count	Percentage
TOP100			
substance		85	83,33 %
fabric		5	4,90 %
knowledge		12	11,76 %
	AVAILABILITY	21	20,59 %
	SYNTHESIS	40	39,22 %
	APPLICATION	30	29,41 %
	ACQUISITION	8	7,84 %
	CONTRIBUTION	3	2,94 %
MID50			
substance		35	70,00 %
fabric		11	22,00 %
knowledge		5	10,00 %
	AVAILABILITY	11	22,00 %
	SYNTHESIS	22	44,00 %
	APPLICATION	14	28,00 %
	ACQUISITION	2	4,00 %
	CONTRIBUTION	2	4,00 %
LAST50			
substance		33	67,35 %
fabric		10	20,41 %
knowledge		8	16,33 %
	AVAILABILITY	8	16,33 %
	SYNTHESIS	24	48,98 %
	APPLICATION	16	32,65 %
	ACQUISITION	2	4,08 %
	CONTRIBUTION	1	2,04 %

The script serves as a conceptual scaffold for interpretation by limiting the range of possibilities of how the meaning of the *A*-constituent and the meaning of the *B*-constituent relate to each other. According to the proposed analysis, it is always the *B*-constituent (i.e.-*stoff*) that contributes the script. Which relationship in particular will be inferred against that background basically depends on the meaning of the *A*-constituent. This is not to deny that the word-

external context in which a *Stoff*-compound appears may render certain otherwise plausible options unlikely.

It should be noted that the script remains stable across all compounds belonging to the *Stoff*-family. It comes in two guises, however. The script may be understood relative to the material domain. To this end we distinguish between substance readings and fabric readings. These we take to be the basic meanings. Apart from that, by way of metaphoric transfer, the script may also be read relative to the immaterial domain, giving rise to what we call knowledge readings.

In view of our findings, it seems tempting to conclude that it is always the head constituent from which the inferred relation is derived in German compounds. As noted in Section 2.3 above, the data discussed in Fanselow (1981b) seems reconcilable with such a view. Since our results are based on a single case study, however, such a conclusion would be premature and, perhaps, simplistic. Whether and how far the results gained for *Stoff*-compounds are generalizable to other German compounds remains to be seen. As a follow-up, it would be straightforward to study German compounds in which *Stoff* forms the *A*-constituent.

Acknowledgements

Earlier versions of this paper have been presented at the workshop "The semantics of derivational morphology: Theory, methods, evidence" at DGFS 2021 and at the conference "Word-Formation Theories VI & Typology and Universals in Word-Formation V", Košice, Slovakia, 2022. We would like to thank the audiences of these events for valuable feedback. We are particularly grateful to Sven Kotowski and Ingo Plag for helpful comments on an earlier draft of this article. Many thanks also to: Swintha Danielsen, Tom Elvins, Volker Gast, Holden Härtl, Matthias Hüning, Martin Schäfer, Hans-Jörg Schmid, Iuliia Shcherbina.

References

- Barsalou, Lawrence W. 1992. Frames, concepts, and conceptual fields. In Kittay, E. A. Lehrer (ed.), *Frames, Fields and Contrasts. New Essays in Semantic and Lexical Organization*, 21–74. Hillsdale: Lawrence Erlbaum Associates.
- Bauer, Laurie. 2017. *Compounds and Compounding*. Cambridge: Cambridge University Press.
- Busse, Dietrich. 2012. *Frame-Semantik. Ein Kompendium*. Berlin, Boston: de Gruyter.
- Busse, Dietrich. 2018. Überlegungen zu einem integrativen Frame-Modell: Elemente, Ebenen, Aspekte. In Ziem, A. L. Inderelst D. Wulf (ed.), *Frames interdisziplinär: Modelle, Anwendungsfelder, Methoden*, 69–92. Düsseldorf: Düsseldorf University Press.
- d'Avis, Franz. 2016a. Einleitung Normalvorstellungen und Sprache. In d'Avis, F. H. Lohnstein (ed.), *Normalität in der Sprache*, 7–36. Hamburg: Buske.
- d'Avis, Franz. 2016b. Exclamatives, Normalcy Conditions and Common Ground. *Revue de Sémantique et Pragmatique* 40, 17–34.
- Fanselow, Gisbert. 1981a. Neues von der Kompositafront oder Zu drei Paradigmata in der Kompositagrammatik. *Studium Linguistik* 11, 43–57.

- Fanselow, Gisbert. 1981b. *Zur Syntax und Semantik der Nominalkomposition*. Tübingen: Max Niemeyer.
- Fillmore, Charles. 1976. Frame semantics and the nature of language. In *Annals of the NY Academy of Sciences: Conf. on the Origin and Development of Language and Speech* 280, 20–32.
- Fillmore, Charles. 1982. Frame semantics and the nature of language. In Linguistic Society of Korea (ed.), *Linguistics in the morning calm*, 111–137. Seoul: Hanshin .
- Fillmore, Charles & Baker, Collin (2009). A Frames Approach to Semantic Analysis. In Heine, B. H. Narrog (ed.), *The Oxford Handbook of Linguistic Analysis*, 313-340. Oxford University Press.
- Fleischer, Wolfgang & Barz, Irmgard. 2012. *Wortbildung der Deutschen Gegenwartssprache*. 4., völlig neu bearbeitete Ausgabe. Berlin, Boston: de Gruyter.
- Hobbs, Jerry & Stickel, Mark & Appelt, Douglas & Martin, Paul. 1993. Interpretation as Abduction. *Artificial Intelligence* 63, 69–142.
- Ingason, Anton & Sigurðsson, Einar. 2020. Attributive compounds. In Lieber, R. (ed.), *The Oxford Research Encyclopedia of Morphology*, Oxford: Oxford University Press.
- Irmer, Matthias & Mueller-Reichau, Olav. 2018. Script Knowledge and the Felicity of Phase Particles in German Adjectival Passives. *Journal of Semantics* 35 (4), 585–637.
- Levi, Judith N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.
- Lieber, Rochelle. 2016. *Compounding in the lexical semantic framework*. In ten Hacken, P. (ed.), *The Semantics of Compounding*, 38–53. Cambridge University Press.
- Löbner, Sebastian. 2015. Functional concepts and frames. In Gamerschlag, T. & Gerland, D. & Osswald, R. & Petersen, W. (ed.), *Meaning, Frames, and Conceptual Representation*, 15–42. Düsseldorf: Düsseldorf University Press.
- Mueller-Reichau, Olav, & Irmer, Matthias. 2023. Knowledge-driven compound interpretation. A corpus study on German complex nouns headed by -stoff [Data set]. To appear in *SKASE Journal of Theoretical Linguistics*. Zenodo. <https://doi.org/10.5281/zenodo.7577641>
- Petersen, Wiebke. 2015. Representation of Concepts as Frames. In Gamerschlag, T. & Gerland, D. & Osswald, R. & Petersen, W. (ed.), *Meaning, Frames, and Conceptual Representation*, 43–69. Düsseldorf: Düsseldorf University Press.
- Schäfer, Martin. 2017. *The semantic transparency of English compound nouns*. Berlin: Language Science Press.
- Schank, Roger & Abelson, Robert. 1975. Scripts, plans, and knowledge. In *Proceedings of the 4th International Joint Conference on Artificial intelligence - Volume 1, IJCAI'75*, 151–157. San Francisco: Morgan Kaufmann.
- Schank, Roger & Abelson, Robert. 1977. *Scripts, Plans, Goals, and Understanding*. Hillsdale: Erlbaum.

Spencer, Andrew. 2011. What's in a Compound? *Journal of Linguistics* 47 (2), 481–507.

*Olav Mueller-Reichau,
University of Leipzig,
Leipzig, Germany
reichau@uni-leipzig.de*

*Matthias Irmer,
OntoChem GmbH,
Halle (Saale), Germany
irmer@conceptmining.de*

*In SKASE Journal of Theoretical Linguistics [online]. 2023, vol. 20, no. 1 [cit. 2023-03-16].
Available on web page <http://www.skase.sk/Volumes/JTL52/02.pdf>. ISSN 1336-782X*