On abstraction in construction grammar

An exercise in methodology

Barend Beekhuizen

On abstraction in construction grammar

An exercise in methodology

THESIS

submitted as partial fulfillment of the degree of

MASTER OF ARTS

in the field of

LINGUISTICS

at the

LEIDEN UNIVERSITY CENTER FOR LINGUISTICS

LEIDEN UNIVERSITY

by

BAREND FRANÇOIS BEEKHUIZEN

Leiden, July 1, 2010

LU CL
Linguistics in Leiden

Universiteit Leiden

Cover image: *The Wedding Dance* (c. 1566) by Pieter Bruegel the Elder.

On abstraction in construction grammar

An exercise in methodology

Supervisor Dr. R.J.U. Boogaart, Leiden University
Co-supervisor Prof. dr. A. Verhagen, Leiden University

Summary

Two important questions in the cognitive sciences are how categories are stored and how people categorize with these knowledge representations. A main dimension in this discussion is the question how abstract a cognitive agent's categorical knowledge is. This dimension forms the subject matter of this thesis. An increasingly popular method of investigating questions like these is developing formal models and evaluating them by means of computational modeling techniques. By simulating learning and categorization processes in a formalized way, we can gain insight in the mechanisms that drive them.

In this thesis, I investigate such a formal methodology for answering the question of abstraction in the grammatical knowledge of a language user. According to constructionist approaches to language, the grammatical patterns employed in linguistic behavior are categories as well. These patterns are stored in a network in which the abstraction of the different items plays an organizational role. Constructionist approaches employ this network in their grammatical analysis, but have no global, non-anecdotal methodology for determining how abstract or concrete the knowledge in it precisely is. More in general, no formalism for categorizing new items using this network has been developed so far.

This thesis aims at working towards such a model. After discussing several alternative categorization models, I apply one of them, viz. Analogical Modeling (Skousen 1989) to grammatical categorization tasks, i.e. ambiguities and form alternations. Analogical Modeling derives from the memory base of exemplars a set of schemas with specified and unspecified values that are used as the basic elements in the categorization task. I claim that the abstraction of a piece of linguistic knowledge (a construction) can be equated with the number of unspecified values of a schema in Analogical Modeling. This way, we can operationalize the notion of abstraction relatively straightforwardly, namely by the number of open slots in the vector of values of a schema. Suppose that the cognitive agent only has a certain set of schemas available, that all fall within a specific range of values for the abstraction parameter (e.g. all schemas have 0 to 4 open slots, but not 5 or 6). We can then evaluate the model given this

set of schemas against observational data, in our case corpus data. Doing so for all possible ranges of abstraction tells us which ranges are optimal for the categorization, i.e. which ranges fit the usage data best.

This model is applied to three categorization tasks for Dutch, using the leaving-one-out cross-validation regime on a set of extracted and annotated corpus data from the Corpus Gesproken Nederlands (Corpus of Spoken Dutch). The data thus consist of relatively spontaneous language behavior. In the first experiment, the dative alternation was studied. The second experiment concerns the disambiguation of transitive relative clauses, which can be ambiguous in Dutch. Finally, the choice of progressive aspect marking, which is a five-way categorization task, was modeled.

The first finding was that for almost all the studied categories, the optimal level of performance was already reached with little abstraction. Using schemas with few open slots, the model was able to generalize towards unseen cases similar to how human agents did. Only little abstraction thus seems necessary. For two categories, viz. the prepositional dative and the progressive with the auxiliary *liggen* 'lie', the model with little abstraction outperformed the model given more abstraction. This means that more abstraction does not always lead to a better command of a category. Too little abstraction hurt the performance as well, as the model was unable to expand the schemas towards new cases.

On the other hand, excluding concrete schemas (with few slots) did not harm the model much either. Only when the used schemas all fell within a highly abstract range, the model was unable to generalize towards new cases properly. On the whole, this means that relatively concrete an relatively abstract models perform equally well. We can consider these cases to be rather uninformative, as they do not allow us to discriminate between any of the ranges of abstraction. Cases that do tell us something about the possible abstraction in the grammatical knowledge of a language user are all minority categories, but are as such all the more interesting. If these minor grammatical patterns would not be categorized correctly, they would disappear quickly from the language, and hence language users must somehow store them in such a way that this knowledge is sufficiently coherent to be generalizable towards new cases. The results for the different categories suggest that too much abstraction and too little concreteness are harmful, as they leads to overgeneralization of the most frequent pattern. To be able to correctly predict the minor patterns, a relatively concrete set of schemas is necessary.

Samenvatting

Twee belangrijke vragen in de cognitiewetenschappen zijn hoe categorieën worden opgeslagen en hoe mensen met deze representaties categorizeren. Een centrale dimensie in dit debat is de vraag hoe abstract de categoriale kennis van een cognitieve agent is. Deze dimensie is het onderwerp van deze scriptie. Steeds populairder in dit type onderzoek is het ontwikkelen van formele modellen. Deze kunnen vervolgens worden geëvalueerd op grote datasets met behulp van computationele modelleertechnieken.

Door het leren en het categorizeren te formalizeren kunnen we inzicht verwerven in de mechanismes achter deze processen.

In deze scriptie onderzoek ik een dergelijke formele methodologie die antwoord kan bieden op de vraag hoe abstract de *talige* kennis van een cognitieve agent is. Volgens de constructionele benaderingen van taal zijn grammaticale patronen ook categorieën. Deze grammaticale patronen worden opgeslagen in een netwerk waarin de abstractie van de patronen een organizerende rol speelt. Hoewel constructionele benaderingen dit netwerk wel gebruiken en beschrijven, is er geen globale, niet op anekdotes beruste methode om te bepalen hoe abstract of concreet de kennis van een taalgebruiker eigenlijk is. In het algemeen ontbreekt een formalisme om nieuwe gevallen te categorizeren met behulp van dit netwerk.

Het doel van deze scriptie is om een aanzet tot de ontwikkeling van een dergelijke model te geven. Nadat ik verschillende categorizatiemodellen heb besproken, pas ik er één, namelijk Analogical Modeling (Skousen 1989), toe op grammaticale categorizatietaken zoals ambiguiteiten en vormalternanties. Analogical Modeling put uit het geheugen, waarin de talige exemplars worden opgeslagen, een verzameling schemas, met open en gespecificeerde waardes. Deze schemas worden gebruikt om nieuwe gevallen te categorizeren. Mijn stelling is dat de abstractie van een element van de taalkennis (een constructie) gelijkgesteld kan worden met het aantal ongespecificeerde waardes in de vector van waardes van een schema. Veronderstel nu dat een cognitieve agent slechts een bepaalde verzameling schema's tot haar beschikking heeft, die gedefinieerd wordt door een bepaald bereik van waardes op die abstractieparameter (bijvoorbeeld: alle schema's hebben 0 tot 4 ongespecificeerde waardes, maar nooit 5 of 6). We kunnen nu evalueren hoe goed een dergelijk model overeenkomt met de geobserveerde gebruiksdata, in ons geval corpus data. Door dit te doen voor alle mogelijke bereiken van abstractie, komen we te weten wat het optimale bereik is voor de categorizatie van nieuwe gevallen, d.w.z.: we komen te weten welk bereik het meest overeenkomt met de gebruiksdata.

Deze procedure wordt toegepast op drie categorizatietaken voor het Nederlands, met behulp van het leaving-one-out cross-validation experimentele regime op geannoteerde corpusdata uit het Corpus Gesproken Nederlands. De data bestaan dus uit relatief spontaan taalgedrag. In het eerste experiment wordt de datiefalternantie bestudeerd. Experiment 2 houdt zich bezig met de desambiguatie van transitieve relatieve bijzinnen, die in sommige gevallen ambigu zijn. Ten slotte modelleer ik de keuze tussen de verschillende progressiefconstructies, een categorizatietaak met vijf uitkomsten.

De eerste bevinding was dat voor bijna alle bestudeerde categorieën het optimale prestatieniveau al bereikt werd met weinig abstractie. Met als kennissysteem slechts schema's met weinig opengelaten waardes was het model in staat om nieuwe gevallen te categorizeren zoals het in de gebruiksdata werd waargenomen. Het lijkt er dus op dat er weinig abstractie nodig is. In het geval van twee categorieën, te weten de prepositionele datief en de progressief met *liggen*, presteerde het model zelfs beter met minder abstractie dan met meer abstractie. Meer abstractie leidt dus niet altijd tot een betere beheersing van een categorie. Te weinig abstractie schaadt de prestatie van het model echter ook, omdat het model in dat geval niet in staat is om te generalizeren naar nieuwe gevallen.

Aan de andere kant schaadt het uitsluiten van concrete schema's (met weinig opengelaten waardes) het model ook niet. Pas als het bereik van abstractie alleen nog die schema's omvat die zeer abstract zijn, is het model niet langer in staat nieuwe gevallen juist te categorizeren. Dit betekent dus dat relatief concrete als relatief abstracte bereiken het vaak even goed doen. Dergelijke gevallen zijn niet erg informatief, omdat ze ons vertellen welk model beter presteert. De gevallen waarin wel een duidelijk optimum te vinden is, zijn allemaal minderheidscategorieën. Als zodanig is het interessant dat juist daar verschillen tussen de bereiken van abstractie aan te treffen zijn. Als deze kleinere grammaticale patronen niet juist gecategorizeerd zouden worden door taalgebruikers, zouden ze erg snel verdwijnen uit de taal en taalgebruikers moeten ze daarom op zo'n manier opslaan dat de kennis voldoende coherent is om uit te breiden naar nieuwe gevallen. De resultaten voor de verschillende categorieën suggereren dat te veel abstractie en te weinig concreetheid vooral schadelijk zijn voor de prestatie, omdat ze tot overgeneralizatie van het meest frequente patroon leiden. Om de kleinere grammaticale patronen correct te voorspellen is een vrij concrete verzameling schema's noodzakelijk.

Acknowledgements

I want to thank Rens Bod, Ronny Boogaart, Pat Clancy, Jack Du Bois, Egbert Fortuijn, Stef Grondelaers, Sandy Thompson and Arie Verhagen for the inspiring classes and discussions that all, in one way or the other, shaped this thesis. I would also like to thank Pieter Rogaar for his help with computer programming.

Ronny Boogaart deserves special mention for being my thesis advisor for two years. He allowed me to undertake several projects, yet always did stretch the important fact that /writing a thesis/ is an accomplishment rather than an activity. All the time, he was an inspiring supervisor and a critical reader of every sketch I wrote during that time. Every time, I left our meetings feeling more motivated than when I entered. Thank you!

An important influence on this thesis is Folgert Karsdorp. Without him having introduced me to Analogical Modeling, LaTeX and the joys of Linux, both form and content of this thesis would be radically different. Furthermore, it was in our discussions that the idea of parametrizing abstraction in categorization models emerged. Although I worked out the details, the cover sheet should essentially have two names on it. I thank you for the continuous support, inspiration and friendship.

Finally, I thank Myrthe for allowing me to occupy the only desk at home and for pulling me away from it when necessary.

Leiden, July 1, 2010,
Barend Beekhuizen

Contents

Chapter 1

Introduction

1.1 On abstraction and grammatical knowledge

When language users verbalize their communicative intentions, a myriad of cognitive processes are called upon to produce the utterance that best conveys these intentions. Among these processes is the ability to assign unknown items to known categories. To express one of the potentially infinite number of communicative intentions that have not been expressed before, language users must be able to employ knowledge about how similar intentions are usually expressed, which involves the use of categorization mechanisms. Major questions in the cognitive sciences are how categories, be they linguistic or not, are represented and how cognitive agents employ these representations. One dimension of this question concerns the degree of abstraction over the input items: do language users categorize with highly abstract patterns, or are only low-level schemas and even concrete memories of experienced language employed in selecting grammatical patterns that convey novel experiences? This dimension forms the subject matter of this study.

On a usage-based constructionist account of language, the grammatical categories are pairings of phonological or structural form and communicative intention. As both knowledge about form and communicative intention (function) are used in the expression of novel intentions or the interpretation of novel utterances, language users must use mechanisms of categorial reasoning to produce and interpret constructions. Because the linguistic constructions are the objects of this categorial reasoning, they can be said to be categories as well and therefore the same questions posed about non-linguistic categories can be asked about constructions too. As Goldberg (2006, 59) states: "grammatical knowledge is knowledge." For partially specified grammatical patterns, we expect to find the same categorization-related phenomena as we find for lexical items, as both are the same kind of mental entities in constructionist theory and for lexical items we do know that they are cognitive categories. Thus, the ditransitive construction, the cleft-clause and the definite NP construction in English are all cases of categories just like 't-shirts', 'things you eat for breakfast' and '1950s formations in football' are.

For lexical classification tasks (e.g. determining whether an object is a cup or a mug), the degree of abstraction necessary to categorize unseen items has been widely investigated, though full agreement has not been reached. Apart from the obvious methods of introspection, observation and empirical experimentation, an important method in studying the mental representation of categories is computational modeling. By defining an explicit, formal model of how categories emerge from the input data, hypotheses about the nature of category storage can be tested for consistency with the observed data.

In the relatively young field of construction grammar, the method of computational modeling has not yet been fully exploited to study abstraction in the knowledge system. Apart from some features indicative of the use of a certain piece of knowledge, such as high token frequency and semantic non-compositionality, there is no general method to determine whether language users classify their communicative intentions as Double-Object Constructions or as Prepositional Dative Constructions based on the most general abstraction over each of them or that they use an array of highly specific patterns. Usage-based construction grammarians seem to prefer a mixture of specific and abstract schemas (Goldberg 2006, 58-59), but are not clear about the conditions for an item to be stored autonomously, nor about the motivation for storing it at a certain level of abstraction. Economy or some principle of non-redundancy cannot form a constraint here, as "concepts and properties in human knowledge are organized with little concern of elegance and parsimony" (Barsalou 1992, 180). Nevertheless, constructionist theories have a conceptual feature that makes a fairly direct investigation of abstraction relatively straightforward, namely the constructicon, or the hierarchical network of constructions and their inheritance relations. In this study, I will exploit this aspect of constructionist theory maximally.

Despite the intuitive appeal, it seems that more knowledge is not always better for a cognitive agent to make the right prediction. Both abstract and concrete knowledge may harm the categorization process, each in different ways. Given that language users are able to categorize using both abstract schemas and highly specific information, we can ask ourselves the question whether they do so for every grammatical categorization problem they encounter. In the case of a construction that differs from similar constructions only in a small and specific set of properties, it seems to provide the maximum effect for the minimum cognitive effort to rely in the categorization process only on the presence or absence of that set and not on irrelevant, overspecific or redundant information. These seemingly irrelevant properties can be harmless, but may also contribute noise to the system, that in turn may cause misclassification if the associations with other, wrong, category labels is strong enough.

A global rule on the other hand, may be so abstract that it has little predictive value in discriminating between the construction at hand and its alternatives. This can happen in cases of patterns with a wide distribution over multiple distinct pockets in the functional space. In some cases, global rules are even impossible, because the category members display only family resemblances (Wittgenstein 1963 (1953), §66-68) to each other, but have no set of features found in all members of the category. Even if a global rule can be extracted, such a rule can possibly even harm the accuracy in deciding between two categories. But how do we know whether the applied rules or schemas are too abstract or too concrete to account for adequate linguistic categorization behavior? To gain overall insight for a certain construction or set of alternating constructions, it is useful to consider a different strategy, namely that of computational modeling.

1.2 Aims of this thesis and overview

The main aim of this thesis is to develop and explore a quantitative methodology for investigating, on the basis of corpus material, how abstract or how specific the constructional knowledge of language users must be for them to engage in succesful linguistic categorization behavior. It does so by building on the ideas developed within the framework of Analogical Modeling (Skousen 1989). Analogical Modeling is an exemplar-based computational model of categorization, in which new cases are compared with experienced cases on the basis of overlapping feature sets of different sizes. Despite being an exemplar model, the use of feature sets that differ in size makes the model highly similar to constructionist approaches in which schemas are posited. This way, AM provides constructional categorization with a formal model.

In AM, the different sizes of the feature sets allow us to investigate the degree of abstraction and specificity in the knowledge system of a language user by parametrizing the range of allowed sizes of feature sets. The question of abstraction then becomes a problem of parameter setting. A source of evidence for preferring one type of setting over the other is derived from the performance of the model with certain settings in terms of its accuracy or predictive power in categorization tasks on the basis of observed corpus data. This ability to predict unseen, novel items adequately is called *generalization*, and is expressed with an accuracy measure. The aim of a modelling study of the type conducted here is to show to what degree a categorization model with only a certain range of constructions (e.g. the concrete ones) fits the usage data better than the model given another range of constructions (e.g. the abstract ones). What is being modeled in this study, is the actual behavior of speakers, as found in corpora of utterances. The model is said to make the right prediction if it predicts the category the speaker employs in that utterance. Convergence to the usage data thus forms the core of the evidence.

Trying to fit the theory to an explicit model and comparing different settings for those models is, or should be, an important scientific virtue in linguistics, more than it is now. In his introduction to the second volume of *The new psychology of language*, Tomasello (2003b, 2) states that one of the hallmarks of Cognitive-Functional linguistics is that it "adopts the categories of traditional Western linguistics only tentatively and provisionally based on their correspondence to the actual patterns of use (. . .)" This vantage point can and should be applied to our ideas about *categories* themselves as well.

The methodology of quantitative, formal modeling forces the researcher to consider the assumptions underlying the model. This often proves insightful, and it can even help one uncover hidden assumptions behind one's theory. Starting from usage-based construction grammar as a cognitive model of linguistic knowledge, some aspects of the theory seem to be unlikely candidates to be used in a formal model, which raises questions about the proper cognitive status of certain key elements of construction grammar, such as 'prototypes' and 'polysemy networks'. In the theoretical chapters, I will present reasons why it is preferable to exclude these phenomena from the ontology of the cognitive model itself, while acknowledging their descriptive value or their existence as higher-level cognitive entities. A major subordinate goal of this thesis is to investigate aspects of the ontology of the constructionist framework critically.

Obviously, the results arrived at in this study are fully dependent on the data and the way these data are encoded. Using spontaneous usage data is therefore a first important prerequisite. In the end, however, the experiments developed only make claims about the coherence or lack of coherence within a data set given a number of features, which I interpret as offline evidence for certain cognitive statuses of an idealized language user. As such, this evidence only provides us with part of the story. We derive claims about category structure by their accordance with data observed in a relatively uncontrolled, unsystematic, yet highly natural way, viz. a corpus. To verify the accuracy of the method developed here, the results will have to be supplemented by the prediction of data from more controlled and systematic online experiments.

This study is organized as follows. In chapter 2, I discuss the main ideas about the representation of categories and usage-based construction grammar and why neither the formal models nor construction grammar currently can provide us with a satisfying method of studying abstraction. Chapter 3 is devoted to a model that combines the most elegant parts of construction grammar with the exactness of a formal model, namely Analogical Modeling. Aspects of this model are parametrized to investigate the amount of abstraction and specificity necessary to predict novel items properly. In a short aside, some aspects of mainstream constructionist theory are criticized. In three corpus experiments (chapters 4 through 6) dealing with grammatical categorization tasks, I will show the possibilities and limitations of this approach by quantitatively and qualitatively evaluating the outcome of the model at different levels of abstraction.

Chapter 2

Linguistic categorization and abstraction

Construction grammarians such as Goldberg (2006) argue that grammatical constructions are categories, learned and employed in the same way any conceptual category, such as ‘chair’, is. If constructions are categories, the same questions about the mental mechanisms behind categorization that are raised in the cognitive sciences, apply to constructions as well. In this thesis, I zero in on one debate, namely that of abstraction in the use of categories. When language users verbalize their communicative intention, they have to use their linguistic knowledge in categorizing this intention as a case where one uses one formal pattern or another to express it. But does this knowledge consist of abstract associations, or of concrete bundles of properties associated with a formal pattern? In other words: how abstract is the linguistic knowledge language users employ?

In this chapter, I discuss ideas about categorization and abstraction, and how these are applied to the usage-based constructionist approach to grammatical knowledge. After this overview, I argue that usage-based construction grammar has no systematic methodology for determining how abstract or specific grammatical knowledge has to be in order for the language user to select the proper constructions and that the existing cognitive models do not provide the conceptual tools for developing a formal categorization model of construction grammar.

2.1 Categorization and abstraction: an overview

2.1.1 From the traditional view to prototype theory

As Lakoff (1987, 6) explains,

> “[in traditional Western thought, categories] were assumed to be abstract containers, with things either inside or outside of the category. Things were assumed to be in the same category if and only if they had certain

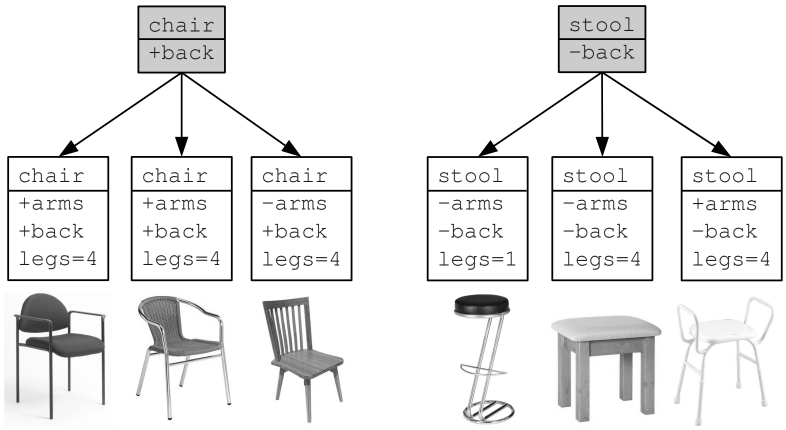


Figure 2.1: A universe consisting of six objects and a subject's categorization according to the classical theory of categorization.

> properties in common. And the properties they had in common were taken as defining the category."

On this view, a category is represented mentally by the set of its defining properties. The set of defining properties of a category is the set of all properties that are shared by all members of the category and that is only instantiated in the members of that category.

Assume a universe consisting of the six objects in figure 2.1. In this universe, a subject perceives four values: a name, whether there is a back or not (± `back`),[1] whether there are arms or not (± `arms`), and the number of legs (`legs =number`).[2] On the traditional view, categories are defined by the set of properties that are both necessary and sufficient, which would in this case amount to the representation of the category 'stool' as ⟨`−back`⟩ and 'chair' as ⟨`+back`⟩. The categorization of new items in this universe proceeds by checking whether the object has a back or not. In the processing of an item, the subject checks whether the necessary and sufficient conditions are met, and then discards the rest of the information.

This theory has three important implications. The first is that all members of a category are equally good members, as they simply share the defining features. Secondly, categories have clear-cut boundaries. Being a chair is a matter of all or nothing: given

[1] In the remainder of this thesis, features and values will be typeset as monospace fonts, sets of values are marked with angular brackets.

[2] Note that these properties are categories themselves. For the ease of exposition, however, we will regard them as atomic percepts.

the set of defining features ⟨+back⟩, a new item will be certainly classified as a chair. The third conclusion, which has not received as much attention as the former two, is that the representation of a category is the most general abstraction that can be made over the experienced instantiations. As this thesis deals with the notion of abstraction, it is important to stress the fact that on the classical account a certain bundle of features is the representation of a category that is used in categorization tasks.

Inspired by the programmatic remarks by Wittgenstein (1963 (1953)), the first departure from the traditional view consisted of a series of experiments conducted by Rosch and colleagues (e.g. Rosch (1978), Rosch & Mervis (1975) and references cited there).[3] An empirical finding disfavoring the traditional model was that on several categorization-related tasks, subjects showed behavior that could not be predicted from a model in which categorization proceeds purely based on the bundle of defining properties of a category. Firstly, not all members are judged equally good members of a category, which would be expected if the representation of a category consisted of a bundle of defining features. Experimental subjects considered robins to be better birds than ostriches, despite the fact that they both have wings and feathers. Secondly, the boundaries with other, neighboring, categories are fuzzy and sometimes indeterminable. For some objects, it is hard to classify them as a stool or a chair, a cup or a mug. Subjects may differ among each other in their categorization behavior in such cases.

The categorization judgments seemed to reflect a different structure of a category than the classical view suggests. A single bundle of necessary and sufficient properties would not do. Rosch, however, did not describe the implementation of such a structure, but rather the effects. She appropriately called them ‘prototype effects’, and did not make any claims about the mechanisms from which these effects emerge. The main prototype effects are the fuzziness of category boundaries, differential judgement on the goodness of membership of categories and the fact that there is often a member of a category that is the ‘best example’ or that has the highest family resemblance, in terms of Wittgenstein (1963 (1953)), to the other instantiations.

After Rosch’s early work, the challenge of implementing the representation of the notion of ‘prototype’, was taken up by many cognitive scientists, leading to a range of interpretations of the concept. In her later work, Rosch emphasizes that prototypes are a “convenient grammatical fiction” representing mainly “judgments of degree of prototypicality” (Rosch 1978, 40). As such, Rosch’s ideas about prototypes do not constitute a new model of category learning, processing or representation, as she states herself as well (Rosch 1978, 36-37). Rather, they form a constraint on models of learning, processing and representation. Whatever model we use, it should be able to account for those phenomena that we call ‘prototype effects’.

The third conclusion drawn from the classical view, namely that category representation is an abstraction over the information of the input, was left relatively un-

[3] Although Posner & Keele’s (1968) work on artificial data sets preceded Rosch’s work with several years.

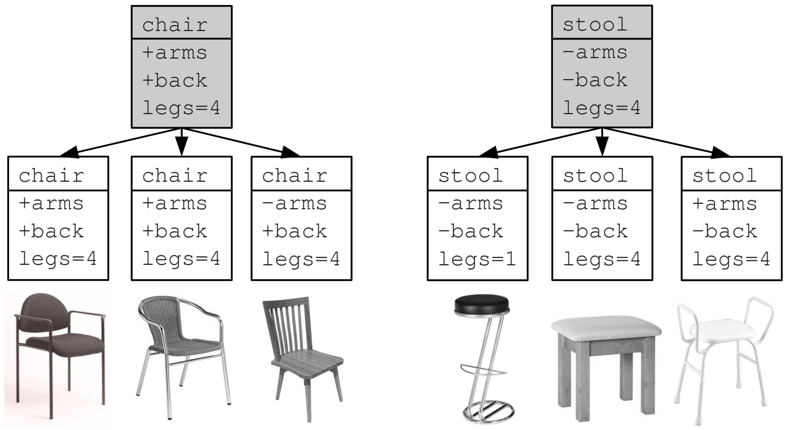


Figure 2.2: A universe consisting of six objects and a subject's categorization according to the abstract-mode prototype theory of categorization. The grey areas are the units used in categorizing new items

challenged by Rosch's findings. Starting from an empiricist point of view, one might still argue that learning and processing categories, and hence the representation of them, proceeds by forming one central summary representation over the input data, as was the case in the classical account. This central abstraction, which is (confusingly enough) sometimes called the 'prototype' in the literature as well, might then be a bundle of properties, but now with the mode value of each property stated in the abstracted prototype. Figure 2.2 represents such a conception of category structure.

Another implementation is that the prototype is a mean-based summary representation that is updated with the processing of each new item. Figure 2.3 displays this. With some properties being categorical, this mode of representation has to involve transforming discrete properties into continuous ones. Having arms is now coded as `arms=1`, where the number stands for the probability of a instance of this category having arms. In the summarization, we can hence find fractions, for instance in the case of the category 'stool'. Instances of this category have a probability $p = 0.33$ of having arms. A new item is categorized by calculating its conceptual distance (the difference on all values) to all of the prototypes and determining which of these it is the most similar to.

What all these models share, is the fact that the representation is an abstraction over the input items by means of clustering. As the clustered representation contains the central tendency of the category, we can call such models *centroid* models.

A radically different implementation is the view on which the prototype is the best member of a category. On this view, the prototype is the set of cases that shares the

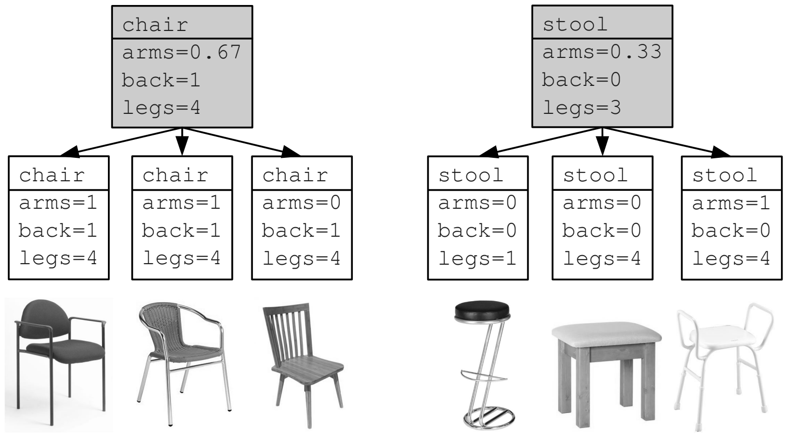


Figure 2.3: A universe consisting of six objects and a subject's categorization according to the abstract-mean prototype theory of categorization. The grey areas are the units used in categorizing new items

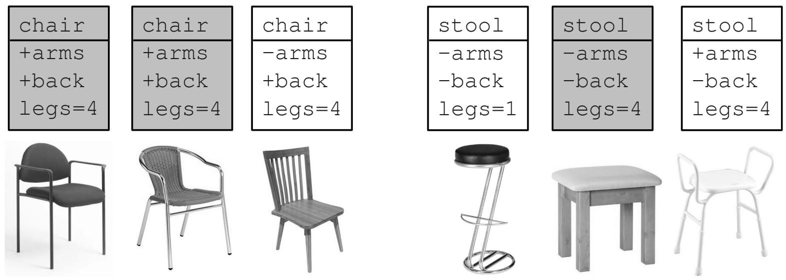


Figure 2.4: A universe consisting of six objects and a subject's categorization according to the best-example prototype theory of categorization. The grey areas are the units used in categorizing new items

most properties with other instantiations of the category and the least with those of other categories. On their own, they can be said to represent the category as a whole metonymically. The best chairs are the leftmost two, and the best stool is the central one (although the leftmost stool can be said to differ more from all of the chairs by virtue of having one leg). Figure 2.4 shows this model of category representation. New items are categorized by comparing them to the grey units, the prototypes, and determining with which one they share the most properties. The importance of this take on category representation lies in the fact that the category is not represented mentally as an *abstraction over* the input items, but as *one of* the items itself. However, this does not make such a model less centroid: the single exemplar representing the category still is a member with properties that reflect the central tendencies of the category.

2.1.2 Formalizing the categorization process: exemplar theory

Rosch left learning theorists with groundbreaking findings and a lot of questions. One of these followed from the initial confusion about the cognitive nature of the notion of prototype. One interpretation of prototypes is that they are mental objects, either as summary representations or as best cases. This interpretation was found in the three modes of category representation discussed in section 2.1.1. Another interpretation of prototypes holds that prototype effects are truly *effects* that may emerge from a representation in which the prototype itself is not a mental entity.

One important family of models that accounts for prototype effects in categorization behavior without reifying prototypes, is that of the exemplar theory. Exemplar theory grew out of the awareness that a cognitively stored central tendency as a representation of a category, as some thought the prototype was, could not account for certain behavioral phenomena, such as the knowledge of the range of values for certain features of categories and inequal membership judgements and categorization times of items that had equal distances to the prototype (Medin & Schaffer 1978). In exemplar theory, a category is not so much represented by an abstraction over all instantiations of a category, but by the memory traces of all of the experienced instantiations of a category themselves. Generalization is achieved by comparing the unseen item with the stored experiences of the instantions of the categories and deciding to what category the new item belongs on the basis of its similarity to the exemplars in memory. Similarity is then calculated as the conceptual distances between the strings of values, i.e. the inverse overlap of the feature sets of the two exemplars.

As Ross & Makin (1999, 215) put it, exemplar theory "seems to take away the 'categoriness' of categories " by no longer conceiving of categories themselves as mental objects but as emergent from distributively stored constellations of memory traces. Abstraction over the input items thus plays no role in the categorization process. Obviously, this is not to say that cognitive agents cannot consciously reason with abstract knowledge, as exemplified in propositions such as 'all birds have wings'. They simply do not use it in subconscious categorization processes.

Among the first operationalizations of the idea of categorization by means of case-based reasoning was the Context Model (Medin & Schaffer 1978). This model was highly succesful in predicting the outcome of online experiments on artificial data. At the same time, the model was able to account conceptually for prototype effects, as the exemplars that display 'average' values are typically the most frequent ones as well and thus weigh heavily in predicting novel items.

In the Context Model, a cognitive agent assigns a category label to a new item by comparing it with all previously categorized members or exemplars of the possible categories.[4] Central to all exemplar models is the idea that the greater the similarity of a new item to a stored exemplar is, the more likely it is that the category label associated with that exemplar is used in categorizing the new item. In simpler terms: the similarity with specific members of categories, not categories themselves, determines classification. No information stored on an abstract, categorial level is thus used in category behavior, although Medin & Schaffer (1978, 211) stress that this does not mean that such information does not exist.

The likelihood of a new item being categorized as a member of a certain category is a function of its similarity to all members of that category. In this respect, it does not differ much from a centroid model, because a centroid representation summarizes the idea of 'the similarity to all members of a category' in one central data point, for instance that of the arithmetic mean. An importance difference between a centroid model and the Context Model, however, is that in the latter, exemplars that are closer to the new item, have more influence on the decision making process. Therefore, local clusters that are relatively dispant from the global central tendency may nevertheless be used to predict novel cases. This is impossible with a centroid model, as all novel items are compared to a single, centroid, data point per category label.

Another implementation of the exemplar model is the theory of Memory-Based Learning (Daelemans & van den Bosch 2005). In this model, the k Nearest Neighbor, or k-NN algorithm is used to categorize new items on the basis of the set of exemplars. Like the Context Model, this model is based on the new item's similarity with stored items rather than its closeness to a centroid abstraction. Essentially, the algorithm selects the set of exemplars (possibly one) that are the most similar to the novel input item and uses the majority outcome of that set as the outcome of the categorization task. Similarity is calculated as a conceptual distance measure and is parametrized by different ways of estimating the weight of the different features and different metrics of computing the amount of overlap between the feature structures of the new item and the exemplar.

In the two exemplar models described here, the ability to generalize to unseen cases is not achieved by abstraction over the cases, but by calculating the similarity between

[4]The notion 'possible' in the phrase *the possible categories* points to a central problem for all categorization approaches, including the one presented in this thesis. Given that a new item has to be categorized in order for an agent to respond appropriately to it (in case it is dangerous, edible, situponable and so forth), there has to be a finite set of categories in the mind of the agent. A major question is where do these categories come from in general. Models presupposing the categories before the first exemplars are encountered, do not address this problem. I believe a viable solution for this problem may lie in techniques akin to unsupervised and incremental Bayesian clustering techniques as employed in Alishahi & Stevenson (2008), where the categories are emergent.

But even granted that the category labels are a priori given to the model, another problem remains, viz. the question where the features come from. The way I present the arms of a chair, is as if they are recognized before the agent has a general notion of chair. A notion of 'arm of a chair', however, is not possible without the notion of 'chair'. The association between the features thus is bidirectional, rather than monodirectional. In this thesis, I will assume monodirectionality between the features and the category labels, although this assumption is in many cases unwarranted.

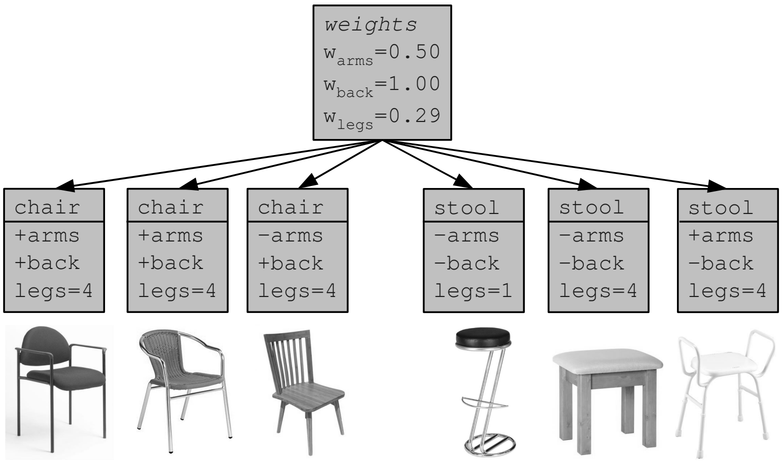


Figure 2.5: A universe consisting of six objects and a subject's categorization according to the Memory-Based Learning theory of categorization. The grey areas are the units used in categorizing new items

known cases and a novel case. The fact that this probabilistic, quantitative reasoning is part of the model's mechanism of making comparisons between the vectors of values that constitute the exemplars, is the reason that the model is able to generalize towards unseen cases without reifying true abstractions. As such, it proves the argument that generalization without abstraction is feasible for a cognitive system.

Obviously, not all features are equally good predictors of the outcome. Having wings is a stronger predictor of something being a bird than the habit of building nests. Because of this, features are weighted for their correlation with the outcomes, as was done as well in the Context Model.

To my mind, this weight extraction is the locus of abstraction in these models. Somehow, the cognitive agent has a stored set of weights for the features, which are extracted from the input items. The agent thus reasons with case-based knowledge (the exemplars) as well as global, abstract categorial knowledge (the weights for that set of category labels). As such, both types of exemplar models are not totally free of abstraction. Figure 2.5 schematically represents the knowledge system of an agent given Memory-Based Learning.

In selecting the one nearest-neighbor, the model is the simplest instantiation of the idea that humans reason by analogy to exemplars: the most similar exemplars determine in which cloud of exemplars the novel case is integrated. The fact that global tendencies of categories are not used, expect for the weight extraction, appears to have no influence on performance: the framework of Memory-Based Learning has achieved very high accuracy scores on many Natural Language Processing applications and has, on a more theoretical level, yielded many insightful analyses of categorization behavior in inflectional morphology (see e.g. Daelemans & van den Bosch 2005).

2.1.3 Formalizing the comparison between models

The ideas of category knowledge as a reified prototype and as a distributed set of exemplars are theoretically incompatible and it is thus desirable to evaluate their performance on the same data set. However, in order to compare the two models, they must be comparable. This means that they will have to have some features in common, or otherwise we could not conclude that the better performance of one model on a data set was due to the fact that it is a prototype model rather than the fact that it used a certain distance measure or that it used some other bias not present in the other model. Only in recent years has this comparability issue gained attention, in the form of the Varying Abstraction Model (Verbeemen et al. 2007, Lee & Vanpaemel 2008).

To a large degree, the debate in the cognitive sciences focused on only two possibilities: either cognitive agents store no information on an abstract level or they represent categories by a fully abstract centroid summary. As the latter implicitly constituted the zero hypothesis, most new models are attempts to account for generalization, the ability to predict new items, without abstraction. Those models thus represent the point of view that a model without any abstraction fits categorization behavior better than does a model with one global, centroid representation per category. The Varying Abstraction Model explicitly compares both models, but also allows for representations in between.

As Lee & Vanpaemel (2008) argue, the two positions that are traditionally identified as the prototype and the exemplar model, can be seen as endpoints on a scale, and can thus both be said to represent the same model but with a different parameter setting. Suppose that we cluster some couples of extremely closely related exemplars together and use the means of those couples as the datapoint with which the new item is categorized (by calculating the similarity between the exemplars and clusters on the one hand and the new item on the other); we then arrive at a model in which there are mainly exemplars, but some local abstractions, which are almost as detailed as are exemplars. If we iterate this clustering process, every time clustering the two most similar datapoints together, we make the level of representation in our model gradually more abstract, and in the end we arrive at a situation in which only the mean values over all exemplars are taken into account, and thus at a completely centroid model. The degree of abstraction in a model, Lee & Vanpaemel (2008) claim, is thus a parameter that is set based on what fits the observed data best.

In our artificial example of the chairs and stools, this amounts to the meta-model displayed in figure 2.6. The parameter, k, represents the degree of abstractness in the category knowledge of a cognitive agent. By testing the performance of the model given different settings of k, we can find in which range of values the model's optimum lies. Suppose that the optimal setting is $k = 4$, then the four centroid clusters on

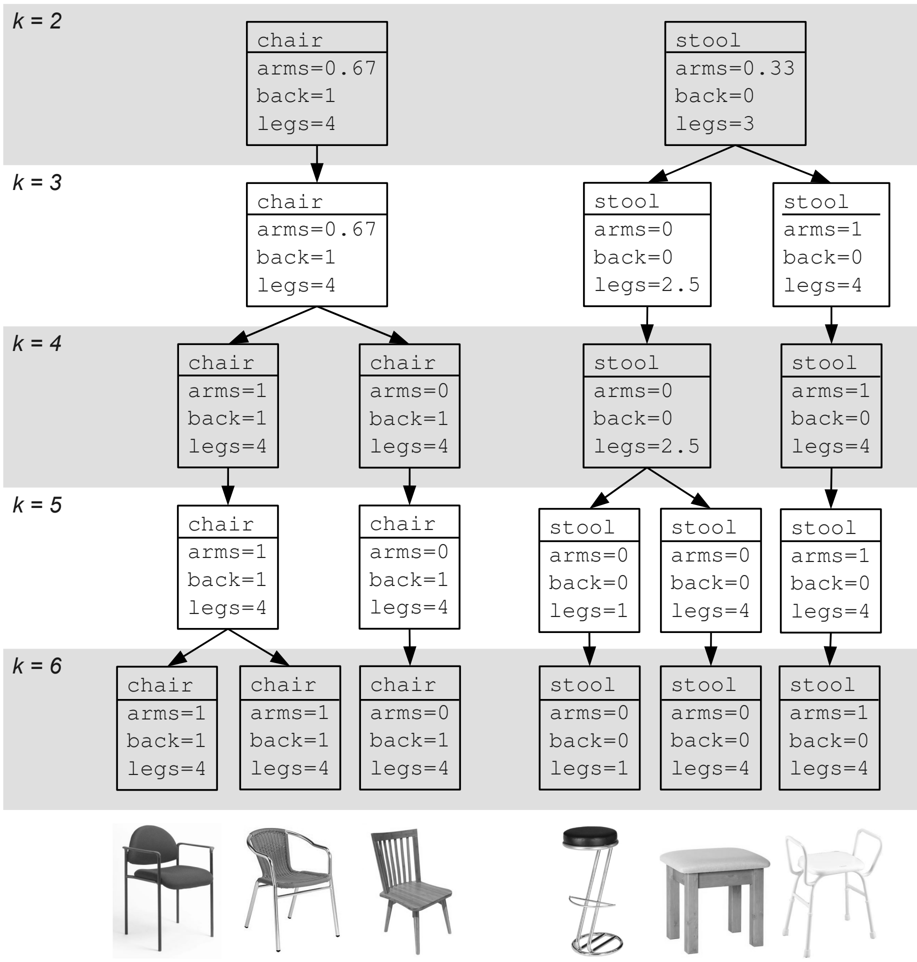


Figure 2.6: A universe consisting of six objects and a subject's categorization according to the context model, as parametrized by the number (k) of clusters. In categorizing using the context model, the subject compares the new item to all memory representations at one level of k.

that layer represent the two categories in the mind of our agent. The new items are compared with these four centroids, using the Context Model, and categorized. Often, it turns out that not the traditional abstract-mean prototype model ($k = c$, where c is the number of categories), or the traditional exemplar model ($k = e$, where e is the number of exemplars) perform optimally, but some value of k in between c and e.

In the Varying Abstraction Model, Bayesian statistical methods are employed to estimate what is the right value for this abstraction parameter, as well as for another parameter, viz. the importance of similarity in the clustering. Crucially, the same weighting and distance metrics are used with all different degrees of abstraction, viz. a generalized version of the Context Model. This way, the different models that are generated by different parameter settings with respect to the level of abstraction, can be compared in an even more controlled way than was done in the earlier studies, where often similarity metrics between exemplar and centroid model differed as well.

Without going into the details of Lee and Vanpaemel's model, the take-home message is that we can gain insight in the abstraction debate by conceptualizing abstraction as a scalar rather than as a categorical value (abstraction or no abstraction), and that by keeping all other metrics for categorizing the new items equal, we can tease out the effect of the presence or absence of abstract representations at different degrees of abstraction. However, the Varying Abstraction Model is not adopted straightforwardly here, as it also has drawbacks for an constructionist account of grammatical knowledge, an issue to which I will come back in section 2.3.2.

2.2 Grammatical patterns as categories

2.2.1 Grammatical patterns as prototype categories

The research on prototype effects had a major influence on the developing functional theories of language, such as Cognitive Grammar (Langacker 1987, Langacker 1991). Rosch found that lexical categories, i.e. the categories represented by content words, display prototype effects, and not much later the realization dawned that also for grammatical morphemes certain uses were more prototypical than others. When constructional approaches emerged, in which the distinction between the syntactic patterns and grammatical morphemes on the one hand and lexical items on the other, began to blur, it could be expected that these grammatical form-meaning pairs were considered to be categories as well, as the case studies in Lakoff (1987) show.

Taylor (1998) explicitly describes how syntactic constructions can be regarded as categories displaying prototype effects. As he argues, a purely formal and categorical description[5] of a construction does not adequately capture the empirical distribution: not all noun phrases and verbs can be used equally well in the ditransitive construction, for instance. Therefore, interactants appeal to the construction's abstract sense, such as the actual transfer meaning associated with the ditransitive construction (cf. Goldberg 1995) to constrain the possible items filling the different subparts of the construction. As some instantiations of a construction are closer to this abstract sense than others, because they involve actual rather than potential transfer or because the theme argument is inanimate, we can expect to find prototype effects in grammatical constructions as well.

Taylor discusses prototype effects that can be found with grammatical constructions. Firstly, different levels of productivity correlate with different levels of centrality. A less prototypical subject, such as the PP *by the campfire* in conversation (2.1),

[5] I believe Taylor means with this 'a description in terms of formal constituents such as NP'.

is not as good a subject as is *he* in sentence (2.2), as *by the campfire* can be used as a subject in only very few argument frames. As Taylor argues, it is restricted to conversations as the one above, where it answers a *where*-question, whereas *he* can function as a subject in many clause types. The notion of subject can thus be said to display prototype effects because of the gradience in the goodness of membership of the instantiations.

(2.1) **A:** Where will we meet?
B: By the campfire is a good spot.

(2.2) He hit me three times.

Secondly, it seems that the boundaries of grammatical constructions are non-rigid and there is leakage from one category to the other. An example is the polysemy within constructions and the issue whether a specific case belongs to the one cluster of uses or to the other. The English past participle has two global meanings,[6] one eventive, meaning that the grammatical subject is getting into the state referred to by the verb (sentence (2.3)), the other stative, meaning that the grammatical subject is in the state referred to by the verb (sentence (2.4)). However, there are cases, such as sentence (2.5), in which the intended meaning of the participle can be stative and eventive at the same time, as some other element (*get*) already bears the eventive meaning.

(2.3) The children were scared by the loud crash.

(2.4) The children were very scared during the earthquake.

(2.5) The children will get scared if you do that.

The intended meaning of this utterance includes both a hypothetical state of the child in the future and the eventive development leading to that state. The use here is not prototypical and can be said to contain functional aspects of both prototypical uses. If we conceptualize those functional aspects as the dimensions of an n-dimensional space, and we give the two prototypical cases each a coordinate in this space, the coordinate of this new item will lie somewhere in between the two prototypical ones.

In regarding a grammatical construction as a category as well, a linguist is commited to take seriously the research that has been done on categorization in general. Whatever the outcome may be, such a venture is an interesting one. When there is correspondence between the predicted way in which we represent conceptual categories such as 'cats' and 'bebop artists' and the predicted way we represent linguistic categories such as noun, the ditransitive and the *way*-construction, we have more evidence to say that large parts of the mental mechanisms behind linguistic behavior depend on domain-general skills.

[6]For an account of the acquisition of these senses, see Israel, Johnson & Brooks (2000), from which the three examples used here were drawn too.

2.2.2 Abstraction in construction grammar

How do people arrive at patterns that display prototype effects and at what level of granularity are the constructions stored? Although none of these questions have been answered with a formal model, usage-based linguists do present clear cases where it is evident that the linguistic knowledge has to be low-level in order to perform properly on linguistic categorization tasks. We can think here of idioms, such as '$\text{NP}_{\text{subject}}$ *kick the bucket*', which is unlikely to be formed on the basis of the abstract transitive construction, as then the idiomatic meaning would be lost. For more 'regular' cases, usage-based linguists do expect that some abstraction is necessary. Two ideas about abstraction stand out: Langacker's theory of schematization and Goldberg's idea about constructional schemas in her usage-based variety of construction grammar.

According to Cognitive Grammar (Langacker 1987), most categories in language are complex categories, meaning that they are not pairings of one invariant form with one invariant meaning: "A speaker's conventional knowledge of [a complex] category cannot be reduced to a single characterization" (Langacker 1987, 370). Rather, Langacker proposes that different senses and forms form a hierarchical network that *as a whole* constitutes the category.

As Langacker (1987, ch. 10) argues, a network consists of nodes, being either schemas or representations of objects, and the relations between these nodes. Given a set of objects with a common category label, say *chair*, the cognitive agent will posit mental schemas that reflect the overlap between the objects. The schemas that emerge are minimally abstract, however. Schemas are abstractions over the features, frames and other semantic structures of the items. They may emerge from the comparison of a new object and a known object, but also from the comparison of a new object and another schema.

Apart from perceived objects and schemas, there are also prototypes in Cognitive Grammar. Langacker claims that the prototype is a privileged node in the network that is used in categorizing new items. How this node gets this status is not entirely clear, but Langacker emphasizes that the assignment of a prototype in a network is a dynamic matter: as experience grows, prototypes may shift. Another important aspect of Cognitive Grammar is that a category may have multiple prototypes. This is the case, for instance, when a category has some metaphorical extension from its base sense. For a linguist, there are two prototypical trees: one being the variety of the plant that occurs most frequent locally, the second a phrase-structure tree.

Importantly, Langacker emphasizes that any depiction of such a network is only a depiction at one moment in time: the knowledge base is updated with every new case of a category that is processed. Unseen cases are categorized by means of the local prototype they resemble most closely and then integrated in the network. This integration includes the emergence of a schematic relation between the new sense and the local prototype, given that the new sense is sufficiently different from that local prototype. If the new sense is almost identical to the local prototype, it is simply included under the prototype. If it is radically different from a prototype, for instance if the domain of experience is different, a metaphorical extension is posited. Langacker claims that only between the new item on the one hand and the local prototype and the immediately coactivated nodes on the other schemas emerge, but not between the new case and all other cases, as there would be too many other cases to draw a comparison with.[7]

Given this theory and our artificial example, a network such as the one depicted in figure 2.7 can be constructed. Suppose the categories are ordered from left to right as they are experienced in time. The middle chair is then categorized based on the first chair, and a schema containing the overlapping structure is posited. The right-most chair is then classified on the basis of this schema, and stored separately as an extension, because it differs on the salient feature `arms`. For the category of stools, things are a bit more complicated. After having classified the second stool as a member of the same category on the basis of a two-out-of-three overlap, the agent forms a schema specifying this overlap. In categorizing the third stool, this schema, however, is not used as the representation of the center stool resembles the new stool to a greater degree. The third stool is thus classified based on the second one and a schema specifying the overlapping semantic structure is posited. As Langacker does not explicitly discusses how prototypes are assigned to the network except for the fact that they are assigned dynamically, the best assumption is that after having seen two stools, the schema ⟨`-arms,-back`⟩ is the prototype, but after classifying the third stool with a directly coactivated node, viz. the representation of the second stool, the prototype shifts to the representation of the second stool, being the stool having the most family resemblances to all other stools.

The constructionist theory of Goldberg (2006) adheres essentially to the view of Langacker (1987), but also incorporates insights from exemplar theory. Goldberg adds, however that abstract schemas[8] over exemplars are stored as well, and in the same way as the exemplars themselves, namely as pairings of an abstract form and an abstract function. Goldberg also emphasizes the dynamics of the representation, especially with respect to language acquisition. Children gradually build up an increasingly abstract network of constructions.

Goldberg provides three reasons for the necessity of abstraction. First, these abstractions are necessary, according to Goldberg, because if they were absent from the knowledge system, linguistic patterns would vary in a more arbitrary manner. Transi-

[7]Extending the idea of language influencing thought to technique influencing thought, I believe we can attribute this argument to the fact that at the time Langacker wrote his book, computers did not have the amount of cpu and RAM they have nowadays. Modern data-oriented theories of knowledge representation, such as Memory-Based Learning (Daelemans & van den Bosch 2005) and Data-Oriented Parsing (Bod 1998) do compare a new item with all previously stored ones and claim the hypothesis space does not get too large because cognitive agents can compute massive amounts of data in parallel. Undoubtably, the idea that human beings can do so is favored more because we know computers can do so.

[8]It is important to keep in mind that *generalization* and *abstraction* are often used interchangeably. Goldberg calls the abstract comparisons that are stored *generalizations*. In order to keep the terminology consistent here, I will refer to them as *abstractions* or *schemas*.

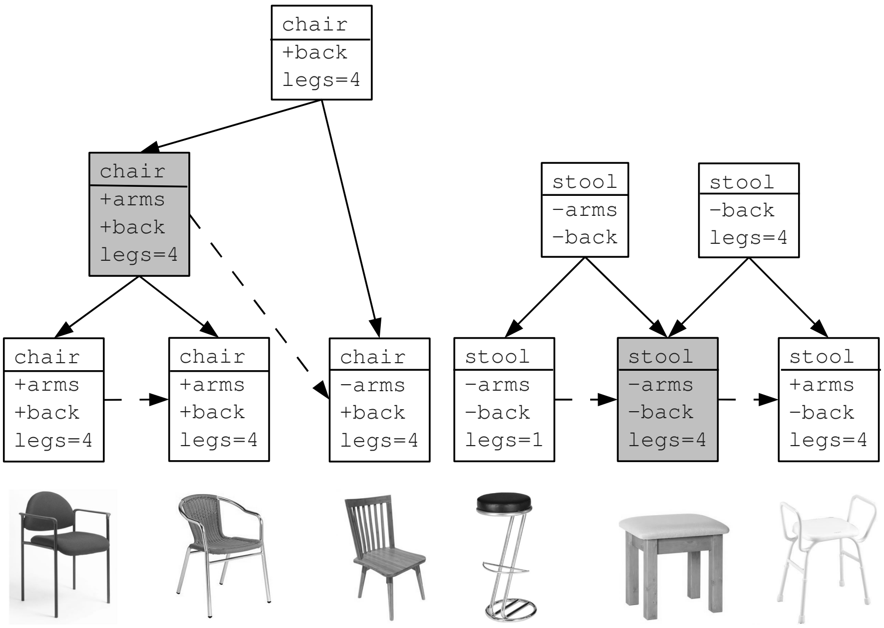


Figure 2.7: A universe consisting of six objects and a subject's categorization according to Cognitive Grammar. The grey areas are the prototypes for each of the category. Dashed lines indicate an 'is categorized based on' relationship, whole lines the 'being an instance of the schema' relationship. The nodes are divided in a top half specifying the phonogical component and a bottom half specifying the semantic component.

tives would be produced in all word orders and varieties of structure thinkable. The abstractions are furthermore needed in predicting the global meaning of novel cases where item-based knowledge would be too fine-grained. Another reason for forming abstractions over argument structure constructions in particular, is that relying on the verb meaning only would decrease the predictability of the overall meaning of the utterance.

Apart from this informal outline of a model, Goldberg discusses a number of effects that are of interest in developing a model. One of these is that a high token frequency of a pattern indicates the storage of that pattern, while open-endedess (a high type-frequency) in one of the slots of the construction indicates that it is stored as a more abstract schema (see also Bybee 2006). Furthermore, she argues that central members of a category play an important role in classifying new items and learning to use a category productively.

The mainstream usage-based vision on abstraction is thus that minimally abstract schemas are extracted when a speaker categorizes a new item on the basis of known items. This contrasts with another take on construction grammar, viz. Construction Grammar (Kay & Fillmore 1999) in which only the most abstract schemas that are not derivable from their component parts are stored. Construction Grammar thus a priori delimits the set of possible constructions used in categorizing new items.

Unlike in Construction Grammar, a usage-based constructionist account allows for low-level, potentially redundant, schemas to become entrenched. Goldberg (1995, 74) calls such a model a *full-entry* model. Despite its name suggesting otherwise, the full-entry model also excludes certain representations, namely abstractions beyond the minimal. Suppose that the lexical construction 'chair' is instantiated by the network displayed in 2.7. The maximally abstract node is then ⟨`+back,legs=4: chair`⟩ that is the minimum overlap between the three exemplars. However, as Langacker (2009) argues, to be able to form schemas that capture the overlap between an exemplar and a new item in the first place, the abstraction allowing for the comparison with the new item must already somehow be in place cognitively. As he calls it, schemas at all levels of abstraction are *immanent* in the processing of exemplars. This abstraction, as represented in a network of schemas or a taxonomic tree, is ultimately only a metaphor for recurrent patterns of neurological activity that leave traces (Langacker 2009, 629).

Taking Langacker's claim seriously and formalizing it, we can say that the abstraction immanent in the processing of an exemplar amounts to extracting from the exemplar all possible subsets of its properties. The abstraction from an exemplar characterized as ⟨`abc`⟩ thus consists of the eight schemas ⟨`abc`⟩, ⟨`ab-`⟩, ⟨`a-c`⟩, ⟨`-bc`⟩, ⟨`a--`⟩, ⟨`-b-`⟩, ⟨`--c`⟩, and ⟨`---`⟩, where '`-`' stands for an unspecified value. Note that the exact string of properties itself, ⟨`abc`⟩, is extracted as well, reflecting Langacker's (2009) claim that the representation of the usage events and the representation of the schemas are of the same cognitive nature. Directing our attention to the network of schemas, we can see that it becomes a much richer representation of categories than the one displayed in 2.7. Nodes such as ⟨`chair,-arms`⟩ and ⟨`chair,+arms,legs=4`⟩ exist as well, but are apparently less interesting in categorization processes.

The set of constructions extracted by the model Langacker informally sets out is thus the collection of the partially ordered sets of the set of values each exemplar contains, as I showed in the extraction of the artificial exemplar ⟨`abc`⟩. The network of the lexical item *chair* is updated with the nodes represented in figure 2.8 when that chair is processed. When the six objects of our artificial universe are processed, the knowledge representation looks like figure 2.10. What we can see here, is that the set of nodes is a superset of the set of nodes specified by the traditional usage-based approaches (full-entry with minimal abstract nodes are extracted), as well as a superset of the set specified by Construction Grammar (only the minimal abstraction covering the maximum amount of cases is extracted, but no redundant ones).

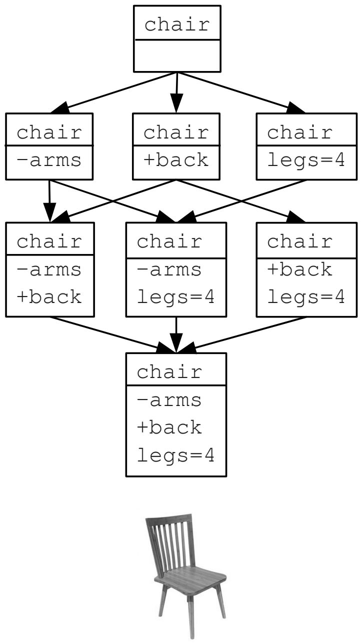


Figure 2.8: The form-meaning pairings extracted in processing an exemplar of the lexical category *chair* according to a formalized version of Langacker (2009). The arrows represent inheritance relations. The nodes are divided in a top half specifying the phonogical component and a bottom half specifying the semantic component.

2.3 Problems and desiderata

2.3.1 Problems with construction grammar as a model

Both Langacker and Goldberg present a model of grammatical organization that is used in categorizing new items using general cognitive mechanisms. As informal expositions, their claims are well-grounded in the research on categorization behavior. One main aspect that is lacking, is that both models do not provide us with an input-output model that can be evaluated not only by single, anecdotal, cases, but also by global tests, evaluating, for instance, the predictive power of the model on the basis of all cases of some construction in a corpus. With an input-output model, I mean an algorithm that can, given certain points of input data, categorize new items of that category by producing a probability distribution over the possible outcome categories. In more theoretical terms: it seems useful to make the procedure of integrating new items and categorizing with this memorized set of exemplars highly explicit and precise. Doing so enables the theory to form predictions of the category of a new item, which can be evaluated against actual performance. Goldberg's nor Langacker's theory provides us with such a model. With construction grammar having the bad name of being occupied solely with exceptions and linguistic nuts, it is important to show how one's model of categorization can account for (approximations of) complete constructional categories as found in corpora.

A reason why both theories could not be directly developed into such a model even if we tried to, is that they lack a formal decision-making component. To be able to claim that a new case is categorized on the basis of some schemas and exemplars that contain a range of features, we have to somehow formalize the comparison between the new case and the mental representation so that the algorithm can decide when to call something a chair and when to call it a stool. The descriptions Goldberg and Langacker give, are reasonably explicit, and can partially be formalized. In a few cases, however, the models remain vague. What, for instance, are the coactivated nodes in Langacker's view on schematization? And how is, given Langackers view of semantics, the distance between the node that forms the local prototype and a new item determined? And when is a schema 'sufficiently frequent' to be stored for Goldberg? When a case is in between two prototypes of different categories, how does the agent make the decision?

The first two desiderata for a model that we can use to investigate abstractness in grammatical categorization are the following.

1 A model should be able to make a prediction on the category outcome of all instantiations of a category. This ability implies that the model has some decision making mechanism to determine which category fits a new case best.

2 A model should have the conceptual mechanisms to compare the semantic and structural properties of two constructions. These include a formalization of the semantic and structural properties and a measure of comparison.

Langacker's (2009) recent exposition of learning mechanisms includes the extraction of all levels of abstraction in the processing of an exemplar. As mentioned earlier, this makes Langacker's recent model a generalization over earlier constructionist approaches, both usage-based and less so. I believe this feature is highly desirable: given a knowledge representation that a priori ranges from the most abstract associations to the most concrete ones, the question how abstract a language user's knowledge is, can be answered in an unbiased way. Models that a priori limit the range of possible abstraction, such as Langacker (1987) and Goldberg (2006), do not allow for the option that in principle very abstract knowledge is used, whereas Kay and Fillmore's Construction Grammar does not allow for the option that very concrete constructions are used, if they are redundant with some more abstract construction. This brings us to the third desideratum.

3 A model should represent all possible abstractions of all exemplars, ranging from the exemplars themselves as minimal abstractions, to associations between the constructional form and an empty set of predicting features as the maximal abstractions.

Finally, Goldberg's argument that languages would differ in more arbitrary ways if no schematization was made, is false in my view, and points to a common reaction to exemplar models. The coherence of a family of constructions may emerge just as well from analogical reasoning over stored exemplars. Using, for instance, a model such as Memory-Based Learning, the coherence is an *effect* of language users trying to use the utterances that are communicatively most appropriate (and hence most similar to known cases), and as such does not have to be built into the cognitive architecture as an abstraction.

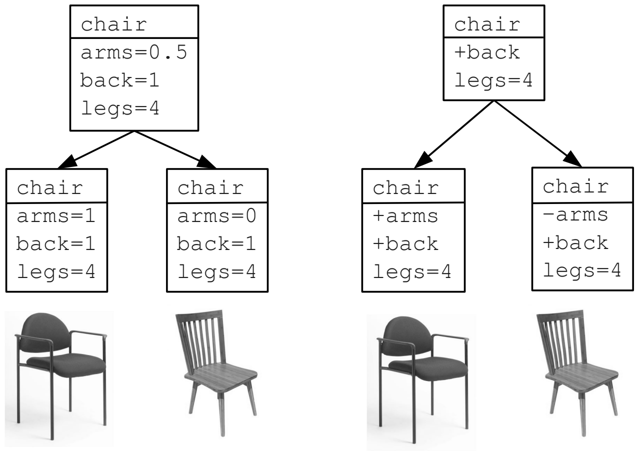
If we think somewhat more about the notion of analogous reasoning or measuring distance between an exemplar and a new case, it seems that these models are not so different after all, as Langacker (2009) notes as well. To compare two items, a cognitive agent must find out what properties are identical and what properties differ. These comparisons, as employed in Memory-Based Learning and the Generalized Context Model, are local abstractions as well. The only non-trivial difference is that in true exemplar models, unlike in constructionist approaches, these local abstractions are never stored but always made on-line. Whether this reflects the actual situation, however, is an empirical question and there are reasons to assume that certain frequently used comparisons are stored in the same way Langackerian schemas are.

Goldberg's argument is thus an instantiation of the more general fear that some property of language (or any phenomenon in general) *would* be left unexplained or *would* be drastically different if the organization was not as the arguer proposes it to be. Often, this kind of argument comes without a thorough investigation of the supposedly false alternative, and it is ironic to see that the generative critics of Goldberg (2006) (Borsley & Newmeyer (2009) and especially Crain, Thornton & Khlentzos (2009)) use the same pattern of reasoning, but then with respect to Goldberg's claim that language users do not form all abstractions linguists see. My criticism, however, should not be taken to mean that I think abstraction is absent from linguistic knowledge. As we will see in the following chapters, I do think language users do make some minimal schematizations over the input.

2.3.2 Problems with models of abstraction as theories

It seems interesting to consider, as Lee & Vanpaemel (2008) do, exemplar and centroid representations as extremes on a scale, rather than as a binary choice. Instead of comparing one exemplar-based model, such as the k-NN model, with another centroid model, such as regression analysis, we can also use just one categorization model with the degree of abstraction being a parameter. The Varying Abstraction Model itself is not suited very well for this venture, for several reasons.

Firstly, it relies strongly on numerical features, whereas language data is often coded most naturally in categorical values. This is a problem that we can solve by



(a) Abstraction by summarization. **(b)** Abstraction by value reduction.

Figure 2.9: Types of abstraction.

transforming categorical features into probability distributions, as was done in the artificial example in 2.6. Certainty of the presence of a feature is then coded with a probability mass of $p = 1$, and certainty of absence with $p = 0$. All values in between represent degrees of certainty. In the clustering, the probabilities of the clustered items are summarized, i.e. with the arithmetic mean, so that we get values like `arms=0.5`, as in figure 2.9a, where the p-value of 0.5 stands for a 50% chance of finding arms if that schema is applicable.

These representations are slightly counterintuitive, but also reflect a different view of abstraction than the one used in construction grammar. Abstractions in the Varying Abstraction Model are summary representations over the datapoints underlying them: they capture the central tendencies of the underlying data points in the arithmetic mean, although other summary representations, such as modes and medians, can be used as well. Given the two exemplars specified by three values, such as the two chairs in 2.9, the abstraction over them, in the form of a centroid cluster, is a summary of the values. In this case, the summary consists of the overlapping values `arms=1` and `legs=4`, and of a probability mass or likelyhood of presence p for the non-overlapping features, namely $p = 0.5$ for the property `arms`. Figure 2.9a shows how the two exemplars are summarized in a more abstract centroid cluster representation.

Summarization, however, is only one possible implementation of the concept of abstraction. Construction grammar employs a different means of abstraction, namely abstraction by property reduction. The idea that constructions form a network of inheritance relations, the constructicon, is grounded in this form of abstraction. Because a parent construction contains less values than its daughter constructions, it can be said to be more abstract. Given our artificial datapoints, a minimal abstraction over these, in the form of a parent schema, would consist of the overlapping values, $\langle$+back$\rangle$ and $\langle$legs= 4$\rangle$, and an open slot for the value of the third feature, meaning that any value is allowed. Figure 2.9b represents this type of abstraction graphically.

With this second type of abstraction reflecting the constructionist ideas about inheritance so closely, it seems that employing a summarization-based model would make the results of such a modeling study hard to interpret. The model needed to investigate abstraction *from a constructionist perspective* will thus have to fulfill the following desideratum:

4 The model should be able to determine which exemplars and schemas will form the basis for the categorization of new items on the basis of categorical reasoning. The mode of abstraction in the model will have to be based on property reduction and inheritance.

The Context Model and Memory-Based Learning do not meet this demand, as they assume only exemplars and no abstract representations on top of them. As models to investigate whether and to what degree language users use abstract knowledge, they are thus not useful.

Another problem with the Varying Abstraction Model is the fact that the abstractions are made on the basis of the outcome categories. The clusters are only formed on the basis of exemplars sharing the category label. As such, the content of clusters at a certain level of abstraction (such as $k = 3$ in figure 2.6) depend fully on the categorization task at hand. Suppose that we have an adjective for the concept $\langle$+arms$\rangle$, say *armish*. It could not rely on the same summarized representations at all levels of k, because certain armish and non-armish exemplars are clustered together at high levels ($k \leq 3$), and this information cannot be straightforwardly used in classifying items as armish or not.

Furthermore, the clustering is done by weighting the relevance of the values in predicting the category labels of the outcomes and by grouping together increasingly relevant differences. The clustering would look different if we were to cluster the dataset into armish and non-armish clusters, as then the weights of the three predictive features, viz $\langle$name,back,legs$\rangle$ woud be different. The knowledge representations thus depend on the outcome rather than on some independent property of the data, whereas we want the extraction of knowledge to be blind for the categorization task it is used for. This leads us to the fifth, and last, desideratum.

5 The model should be extracting knowledge in such a way that we can make predictions about the values of any feature on the basis of the other values.

The models using the property reduction mode of abstraction, such as Goldbergian construction grammar and Cognitive Grammar, are such models. Memory-Based learning is capable of these actions as well, but then the model would have to recalculate the weights for every new feature that is predicted.

A model meeting these five desiderata is Analogical Modeling (Skousen 1989). In this model, a whole network, as the one depicted in 2.10, is used in categorizing a new item. By operationalizing the abstraction of a node in this network as the number of values left unspecified, we can parametrize the range of abstraction as the allowed number of values in any node. The next chapter discusses Analogical Modeling in detail.

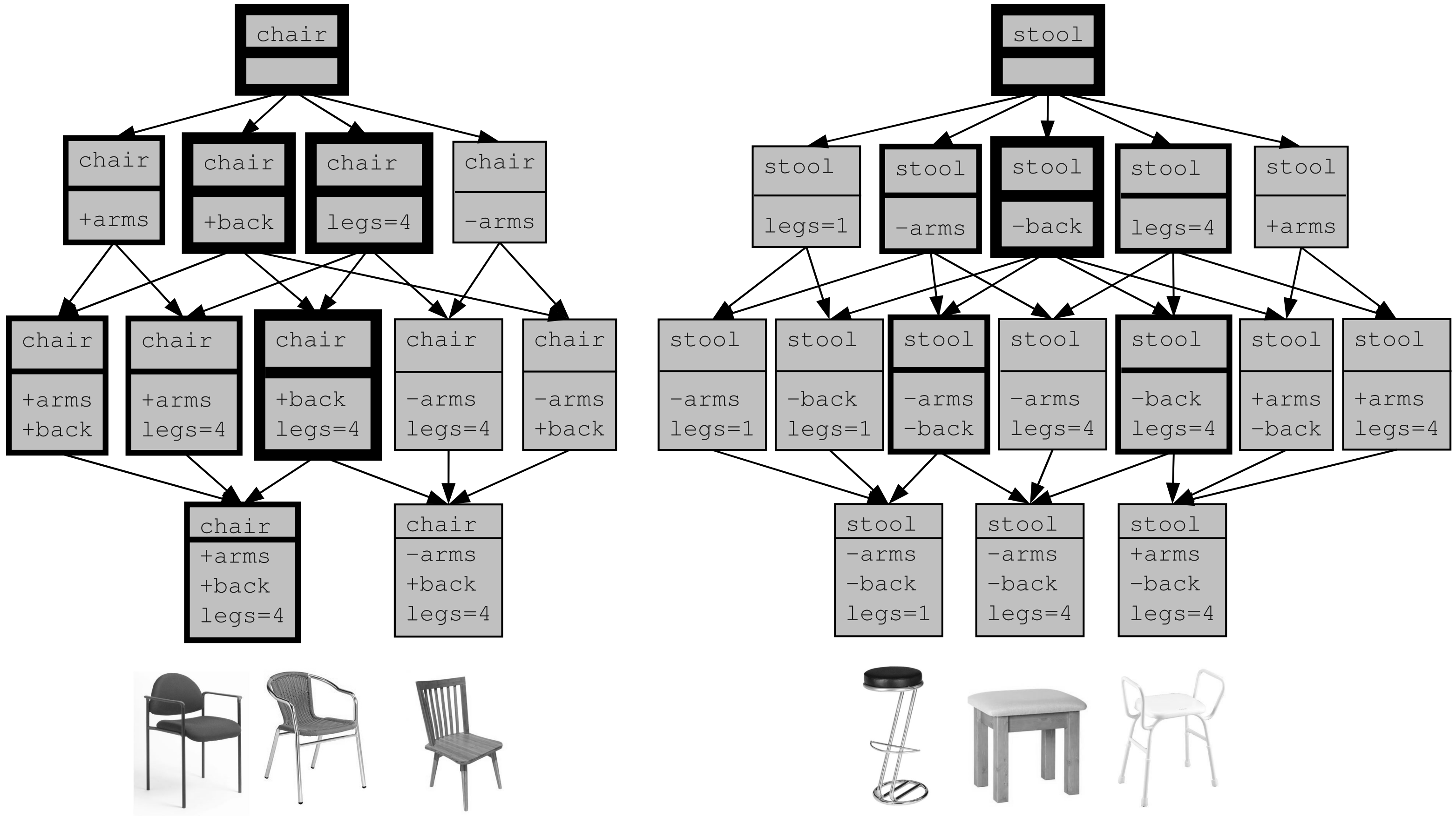


Figure 2.10: The full network representation of the lexical categories *stool* and *chair* , in a formalized version of Langacker (2009). The arrows indicate which schema is schematic over which schema. The grey areas are used in categorization. The thickness of the lines of the boxes represents the frequency of the schema. The nodes are divided in a top half specifying the phonogical component and a bottom half specifying the semantic component.

Chapter 3

A method to investigate abstraction

3.1 Analogical Modeling and abstraction

Given the desiderata of a model to investigate abstraction in grammatical knowledge, I present Analogical Modeling as a likely candidate, and explain to what extent it fits the conceptual foundations of constructionist approaches to language.

3.1.1 The model

Developed in the 1980s, Analogical Modeling (Skousen 1989, Skousen, Lonsdale & Parkinson 2002) is grounded in the assumption that in categorization, cognitive agents simultaneously employ comparisons or analogies between exemplars at different levels of abstraction. Categorization, then, is driven by the overlap of properties of the new item with sets of exemplars in memory that form reliable predictors of a category label. As Analogical Modeling (AM) takes learning to be based on concrete instantiations of a phenomenon, it is best illustrated by an example itself. To show how the idea of analogy is implemented in the model, it is furthermore useful to take an example from a domain where linguists are used to see analogy: morphology.

The Dutch diminutive suffix displays several allomorphic forms. To keep the example feasible, we focus only on a subset, namely the allomorphs used for stems ending in liquids and nasals. The three allomorphs applicable to this subset are:

- /-əcə/, written as [-etje],
- /-cə/, written as [-tje],
- /-pjə/, written as [-pje]

Let us call these categories `etje`, `tje` and `pje` respectively. Suppose that we want to predict the form of the affix for a novel word, say *baam*. Which one of these three allomorphs applies to our new word?

				variables		
number	orthographic	meaning	allomorph	onset	nucleus	coda
1	*raam*	window	pje	r	A	m
2	*boom*	tree	pje	b	O	m
3	*zoom*	hem	pje	z	O	m
4	*naam*	name	pje	n	A	m
5	*wal*	wall	etje	w	a	l
6	*bar*	bar	etje	b	a	r
7	*kom*	bowl	etje	k	o	m
8	*lam*	lamb	etje	l	a	m
9	*bal*	ball	etje	b	a	l
10	*rol*	role	etje	r	o	l
11	*hol*	hole	etje	h	o	l
12	*bom*	bomb	etje	b	o	m
13	*haal*	stroke	tje	h	A	l
14	*laan*	lane	tje	l	A	n
15	*koor*	choir	tje	k	O	r
16	*kool*	cabbage	tje	k	O	l
17	*loon*	wage	tje	l	O	n
18	*taal*	language	tje	t	A	l
19	*paar*	pair	tje	p	A	r

Table 3.1: The nineteen diminutive exemplars stored by an artificial agent.

Suppose that our cognitive agent has heard and stored the nineteen diminutives in table 3.1. The agent has split the sound structure of the noun stems into the `onset`, the `nucleus` and the `coda`. Now, the outcome of the new item is determined by comparing its values on these three phonological properties with the stored exemplars. Because matching sets of values are the means of comparing the new item with exemplars, this model can be said to work on the basis of analogy. We will shortly see how this works.

Central to the determination of the analogues or predictors for the new item in AM is the notion of *context*. A context is a vector of values that defines a set of exemplars in the agent's memory. The values of the new exemplar itself form the *given context*. This given context is the starting point for the search query for analogues. Our new word *baam* thus makes up the given context ⟨`bAm`⟩, short for ⟨`onset = b, nucleus = A, coda = m`⟩.[1] Note that it does not matter whether the categorization model is

[1]For this example and for the case studies, I will use the following convention. As contextual de-

applied to a symbolic categorization task, such as a lexical or constructional choice, or a phonological one. As Langacker (1987, 388-396) argues, grammatical alternatives, lexical choice, the choice of allomorphs and even phonological variation should be accounted for with the same categorization mechanisms, as categorization is a domain-general cognitive ability.

The prediction in AM is done on the basis of the overlap of values between the new item and the exemplars. A context specifying which features are identical, is called a *supracontext*. These supracontexts form the cornerstone of Analogical Modeling, and in a way replace the traditional linguistic mechanism of four-part analogy in linguistics. Furthermore, as I will show shortly, they display organizational properties similar to those of constructions in a network. Before we can discuss how supracontextual comparisons enable the agent to make predictions, we will have to deal with another type of context necessary to determine the usefulness of supracontexts.

In looking for analogues, a set of all possible *subcontexts* divides the set of exemplars into different classes according to their similarity to the novel item. A subcontext is a context, thus a string of values, that specifies the matches and mismatches of values between the new item and exemplars in the memory of the agent. Identity in values is symbolized with the simple statement of the value, whereas difference is depicted with the statement of the value with a bar above it. In the case of our example, the subcontext $\langle \mathtt{b}\bar{\mathtt{a}}\bar{\mathtt{m}} \rangle$ defines all exemplars that do *not* have the values `nucleus=a` and `coda=m` and that do have the value `onset=b`. For our data set, this amounts to the exemplars *bal* and *bar*. Given that exemplars are stored as a vector of n features, there are 2^n subcontexts specified by the given context.

All of the items in the exemplar set can then be assigned to one of these subcontexts. Figure 3.1 presents all subcontexts of the given contexts ⟨`bAm`⟩ and table 3.2 shows the sets of exemplars defined by the subcontexts. Obviously, contexts can specify empty sets of exemplars if there is no exemplar in the exemplar set that satisfies the constraints imposed by the subcontext.

A first step in determining which sets of exemplars form reliable analogues for the new item, we start by calculating the *disagreement* within each subcontext. The

scriptions are sets of values they are presented in a monospace font and between angular brackets. For the sake of brevity, only the values are presented. Because of this, the order of the values is significant and reflects the order of the values, as given for each dataset. The category labels are often left out, but can be given after a colon at the end of the string of values but within the brackets.

In the case of the diminutive dataset, the three values are, from left to right

- the value of the ONSET feature. Every grapheme stands for its regularly corresponding Dutch phoneme.
- the value of the NUCLEUS feature, with capital letters standing for the tense variant of that vowel and the regular ones for the lax variant of that vowel.
- the value of the CODA feature. Every grapheme stands for its regularly corresponding Dutch phoneme.

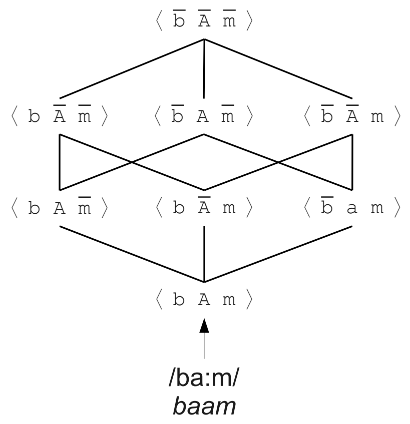


Figure 3.1: The subcontexts of the given context ⟨bAm⟩, presented as a lattice.

disagreement $d(c)$ of a context c, is calculated by comparing the outcome labels of all exemplars specified by c. All outcome labels of exemplars $e_{1...n} \in c$ are compared with all outcome labels of exemplars $e_{1...n} \in c$, hence including a comparison with themselves. If the category labels of two exemplars are not identical, they can be said to disagree with each other. each disagreement from an exemplar $e_i \in c$ with an exemplar $e_j \in c$, the disagreement of the context $d(c)$ is incremented with one.

For a subcontext c_i in which all exemplars have the same outcome label, the disagreement is $d(c_i) = 0$. Subcontext ⟨b̄Ām⟩, however, has exemplars with two different outcomes, and therefore there is disagreement. Table 3.3 shows how we can calculate this disagreement. For this subcontext, the disagreement $d = 4$. Another important fact that we will use later on, is that we can also calculate the *agreement*. The agreement within ⟨b̄Ām⟩ for the outcome `etje` can calculated by counting the agreements between exemplars displaying to the outcome `etje`, and thus is 4. The agreement within this subcontext for `pje` is 1, as there is only one exemplar, pointing to itself. This number of disagreements per subcontext of ⟨bAm⟩ is listed in the right column of table 3.2.

In AM, we predict the outcome category of a new item not directly on the subcontexts. The comparison is not based on the statement of presence or absence of a value, but on the basis of the presence or the unmarkedness of a value. A subset of the total set of values is used as a predictor, ignoring the values on the other features. Contexts that determine values and 'wildcards' for certain values are called *supracontexts*. Because in supracontexts schemas are left unspecified or open, they can be truly said to be schemas.

In the supracontexts, the lack of specification of a value is denoted with the sign

subcontext	exemplars (orthographical form)	$d(sub)$
$\langle\texttt{bAm}\rangle$	$\emptyset$	0
$\langle\bar{\texttt{b}}\texttt{Am}\rangle$	*raam*$_{\texttt{pje}}$, *naam*$_{\texttt{pje}}$	0
$\langle\texttt{b}\bar{\texttt{A}}\texttt{m}\rangle$	*boom*$_{\texttt{pje}}$, *bom*$_{\texttt{etje}}$	2
$\langle\texttt{bA}\bar{\texttt{m}}\rangle$	$\emptyset$	0
$\langle\bar{\texttt{b}}\bar{\texttt{A}}\texttt{m}\rangle$	*zoom*$_{\texttt{pje}}$, *kom*$_{\texttt{etje}}$, *lam*$_{\texttt{etje}}$	4
$\langle\bar{\texttt{b}}\texttt{A}\bar{\texttt{m}}\rangle$	*haal*$_{\texttt{tje}}$, *laan*$_{\texttt{tje}}$, *taal*$_{\texttt{tje}}$, *paar*$_{\texttt{tje}}$	0
$\langle\texttt{b}\bar{\texttt{A}}\bar{\texttt{m}}\rangle$	*bar*$_{\texttt{etje}}$, *bal*$_{\texttt{etje}}$	0
$\langle\bar{\texttt{b}}\bar{\texttt{A}}\bar{\texttt{m}}\rangle$	*wal*$_{\texttt{etje}}$, *rol*$_{\texttt{etje}}$, *hol*$_{\texttt{etje}}$, *koor*$_{\texttt{tje}}$, *kool*$_{\texttt{tje}}$, *loon*$_{\texttt{tje}}$	18

Table 3.2: The subcontexts of the given context $\langle\texttt{bAm}\rangle$ and the exemplars defined by it. The allomorphic categories are presented in subscript after the orthographical form of the exemplars in the second column.

	$\langle\texttt{zOm:pje}\rangle$	$\langle\texttt{kom:etje}\rangle$	$\langle\texttt{lam:etje}\rangle$
$\langle\texttt{zOm:pje}\rangle$	agree	disagree	disagree
$\langle\texttt{kom:etje}\rangle$	disagree	agree	agree
$\langle\texttt{lam:etje}\rangle$	disagree	agree	agree

Table 3.3: Agreements and disagreements between the exemplars in the subcontext $\langle\bar{\texttt{b}}\bar{\texttt{A}}\texttt{m}\rangle$.

'–'. For our case, the supracontext $\langle\texttt{-A-}\rangle$ thus specifies that the value `nucleus = A` is present, and that the other values can be anything. This supracontext therefore effectively encompasses all exemplars also contained by the subcontexts $\langle\texttt{bAm}\rangle$, $\langle\texttt{bA}\bar{\texttt{m}}\rangle$, $\langle\bar{\texttt{b}}\texttt{Am}\rangle$ and $\langle\bar{\texttt{b}}\texttt{A}\bar{\texttt{m}}\rangle$. These subcontexts can thus be said to be subcontexts *of* the supracontext $\langle\texttt{-A-}\rangle$. Given n features, a given context has 2^n supracontexts. The eight supracontexts of the given context $\langle\texttt{bAm}\rangle$ are $\langle\texttt{bAm}\rangle$ itself, $\langle\texttt{bA-}\rangle$, $\langle\texttt{b-m}\rangle$, $\langle\texttt{-Am}\rangle$, $\langle\texttt{b--}\rangle$, $\langle\texttt{-A -}\rangle$, $\langle\texttt{--m}\rangle$, and $\langle\texttt{---}\rangle$. From a rule-based perspective, these supracontexts can be said to be some sort of very detailed rules, stating something like 'if the word has a tense /a/ as its nucleus and an /m/ as its coda, and any onset, produce the /-pjə/ allomorph'.

The supracontexts are the units of comparison. The exemplars singled out by the supracontext are the potential analogues for the determination of the outcome label of the new item. As the supracontexts are hierarchically ordered, with more abstract ones (such as $\langle\texttt{b--}\rangle$) encompassing all exemplars and possibly more than more specific ones (such as $\langle\texttt{bA-}\rangle$ and $\langle\texttt{b-m}\rangle$), exemplars that are very similar to the given context are present in many supracontexts. Because of this, they have more influence on the category outcome of the new item.

However, not all supracontexts can be used in predicting the outcome of the new item. Only if the supracontext is a reliable predictor of the outcome it can be used in the prediction. The exemplars of a supracontext are used in the prediction given that that supracontext is *homogeneous*. A supracontext is said to be homogeneous if and only if the amount of disagreements between the items in the supracontext is not greater that the summed amount of disagreements of all the subcontexts of that supercontext. The rationale behind this is that the supracontext only is a reliable predictor if all of the supracontext's subcontexts behave exactly like the supracontext. If the different subcontexts point at different categorizations or display different tendencies, the supracontext is not coherent enough in its predictive power and cannot be said to be a reliable predictor of the outcome. More specific supracontexts can, in that case, still be homogeneous, but the outcomes of the present supracontext are too diverse.

The constraint of homogeneity effectively makes three types of supracontexts homogeneous:

- Supracontexts containing only exemplars with one outcome type.
- Supracontexts containing no exemplars.
- Supracontexts containing one subcontext with multiple outcome types and besides that subcontext only empty subcontexts.

The homogeneity of the first two is easily understood. Given that all exemplars specified by a supracontext have the same outcome, there is no disagreement in the supracontext, nor in any of the subcontexts belonging to that supracontext to sum up. The same holds for empty supracontexts: there is no disagreement in the supracontext, and hence none in the subcontexts of the supracontext either. As there are no exemplars in these classes that could form analogues for the new items, these supracontexts are not very interesting in making predictions. In the third case, the supracontext is said to be a reliable predictor because the subcontext with multiple outcome types behaves exactly like the supracontext, as the other subcontexts are empty and thus do not increase the summed uncertainty of the subcontexts.

In the case of our example, the analysis is given in the following tables. Table 3.4 contains the supracontexts of the given context $\langle \texttt{bAm} \rangle$, the exemplars that match them and the supracontextual disagreement, or the disagreement among the exemplars that match the supracontextual string.

With this information, we can determine which supracontexts form reliable predictors of the outcome of the lexical choice for the object $\langle \texttt{bAm} \rangle$. As said, this is done by comparing the summed amount of the disagreement of a supracontext's subcontexts to the disagreements among the exemplars of that supracontext itself. Table 3.5 shows this comparison. Empty supracontexts can be left out, as they contain no items that can predict the outcome of our new exemplar. Then, supracontextual schema's for which all exemplars underlying it point in the same direction (that are *deterministic*) are homogeneous too. In our case, the only non-empty deterministic supracontext is $\langle \texttt{-Am} \rangle$, the schema subsuming *raam* and *naam*. Then there is another case in which there is disagreement, but for which the supracontextual disagreement is not higher than the summed subcontextual disagreement. This is the supracontext ⟨`b-m`⟩, the schema subsuming *bom* and *boom*. Here we see the motivation why we would want such supracontexts to be homogeneous: the supracontextual schema ⟨`b-m`⟩ might show diverse outcomes, but it is the best predictor with the `onset = b`, for instance, because the other supracontexts, ⟨`bA-`⟩ and ⟨`b--`⟩, provide us with an empty set and an even less coherent schema.

This analysis leaves us with two non-empty homogeneous supracontexts for the given context ⟨`bAm`⟩, as is shown in table 3.6. The exemplars in these supracontexts are then used to predict the outcome and can thus be said to be the analogues of *baam*. This prediction can be done in two ways: by means of the number of exemplars per supracontext or by means of the number of agreements per supracontext. In the first method, we simply score the total number of times each outcome occurs with the exemplars of each supracontext. This yields for the present example a score of 3 for the outcome `pje` and 1 for `etje`. By majority vote, the selected lexical label for this object would be `pje`, thus: *baampje* (/baːmpjə/). We can also consider this score as a probability distribution and select an item by means of random selection, the outcome `etje` would have a probability of $p = \frac{1}{4}$ against a probability $p = \frac{3}{4}$ for the outcome `pje`. In this thesis, only the majority vote method will be considered.

The second way of calculating the outcome, is by counting the number of agreements among the exemplars in each homogeneous supracontext rather than the number of exemplars. This means effectively that the number of votes for each category label per supracontext is squared and added to the total. For our case, this would make the score for `pje` to be 5 against 1 for the category label `etje`. The selected item by majority vote would with this metric also be `pje`. The effect of this squared metric is that supracontexts specifying more exemplars will become more influential in the analysis. This metric, emphasizing the effect of 'gangs' of exemplars will be used in the following experiments as well.

3.1.2 Introducing abstraction

Skousen's model lends itself well for operationalizing the degree of abstraction in the grammatical knowledge of a speaker. We can think of the different levels of abstraction in the supracontexts as constructions at different hierarchical levels of the constructicon. After all, both are schemas licensed by a set of exemplars used in categorizing new cases. Given this conceptual paralellism, which is worked out in more detail in section 3.2, we can explore the degree of abstraction in the constructicon by looking at the use of more or less abstract supracontexts in classification tasks. To operationalize the notion of abstraction, we simply use the number of unspecified values of a supracontext. Thus, abstraction is the degree to which a supracontext is schematic. The more values a supracontext specifies, the less abstract, or more concrete it is. A supracontext thus has a certain value that reflects its degree of abstraction. This value will be

supracontext	exemplars	$d(supra)$
$\langle\mathtt{bAm}$	$\emptyset$	0
$\langle\mathtt{-Am}\rangle$	*raam*$_{\mathtt{pje}}$, *naam*$_{\mathtt{pje}}$	0
$\langle\mathtt{b-m}\rangle$	*boom*$_{\mathtt{pje}}$, *bom*$_{\mathtt{etje}}$	2
$\langle\mathtt{bA-}\rangle$	$\emptyset$	0
$\langle\mathtt{b--}\rangle$	*boom*$_{\mathtt{pje}}$, *bom*$_{\mathtt{etje}}$, *bar*$_{\mathtt{etje}}$, *bal*$_{\mathtt{etje}}$	6
$\langle\mathtt{-A-}\rangle$	*raam*$_{\mathtt{pje}}$, *naam*$_{\mathtt{pje}}$, *haal*$_{\mathtt{tje}}$, *laan*$_{\mathtt{tje}}$, *taal*$_{\mathtt{tje}}$, *paar*$_{\mathtt{tje}}$	16
$\langle\mathtt{--m}\rangle$	*raam*$_{\mathtt{pje}}$, *naam*$_{\mathtt{pje}}$, *boom*$_{\mathtt{pje}}$, *bom*$_{\mathtt{etje}}$, *zoom*$_{\mathtt{pje}}$, *kom*$_{\mathtt{etje}}$, *lam*$_{\mathtt{etje}}$	24
$\langle\mathtt{---}\rangle$	(all exemplars)	232

Table 3.4: Disagreement in the supracontexts of the given context $\langle\mathtt{bAm}\rangle$. The allomorphy categories are presented in subscript after the orthographical form of the exemplars.

supracontext	subcontexts	$d(supra)$	$\sum d(sub_{i...n})$
$\langle\mathtt{bAm}\rangle$	$\langle\mathtt{bAm}\rangle$	0	0
$\langle\mathtt{-Am}\rangle$	$\langle\mathtt{bAm}\rangle$, $\langle\mathtt{\bar{b}Am}\rangle$	0	0
$\langle\mathtt{b-m}\rangle$	$\langle\mathtt{bAm}\rangle$, $\langle\mathtt{b\bar{A}m}\rangle$	2	2
$\langle\mathtt{bA-}\rangle$	$\langle\mathtt{bAm}\rangle$, $\langle\mathtt{bA\bar{m}}\rangle$	0	0
$\langle\mathtt{b--}\rangle$	$\langle\mathtt{bAm}\rangle$, $\langle\mathtt{b\bar{A}m}\rangle$, $\langle\mathtt{bA\bar{m}}\rangle$, $\langle\mathtt{b\bar{A}\bar{m}}\rangle$	6	2
$\langle\mathtt{-A-}\rangle$	$\langle\mathtt{bAm}\rangle$, $\langle\mathtt{\bar{b}Am}\rangle$, $\langle\mathtt{bA\bar{m}}\rangle$, $\langle\mathtt{\bar{b}A\bar{m}}\rangle$	16	0
$\langle\mathtt{--m}\rangle$	$\langle\mathtt{bAm}\rangle$, $\langle\mathtt{\bar{b}Am}\rangle$, $\langle\mathtt{b\bar{A}m}\rangle$, $\langle\mathtt{\bar{b}\bar{A}m}\rangle$	24	6
$\langle\mathtt{---}\rangle$	$\langle\mathtt{bAm}\rangle$, $\langle\mathtt{\bar{b}Am}\rangle$, $\langle\mathtt{b\bar{A}m}\rangle$, $\langle\mathtt{bA\bar{m}}\rangle$ $\langle\mathtt{\bar{b}\bar{A}m}\rangle$, $\langle\mathtt{\bar{b}A\bar{m}}\rangle$, $\langle\mathtt{b\bar{A}\bar{m}}\rangle$, $\langle\mathtt{\bar{b}\bar{A}\bar{m}}\rangle$	232	24

Table 3.5: Supracontextual analysis of the given context $\langle\mathtt{bAm}\rangle$.

	exemplars per outcome			agreements per outcome		
supracontext	`etje`	`tje`	`pje`	`etje`	`tje`	`pje`
$\langle\mathtt{-Am}\rangle$	0	0	2	1	0	4
$\langle\mathtt{b-M}\rangle$	1	0	1	1	0	1
total	1	0	3	1	0	5

Table 3.6: Homogeneous non-empty supracontexts of the given context $\langle\mathtt{bAm}\rangle$ and their outcomes.

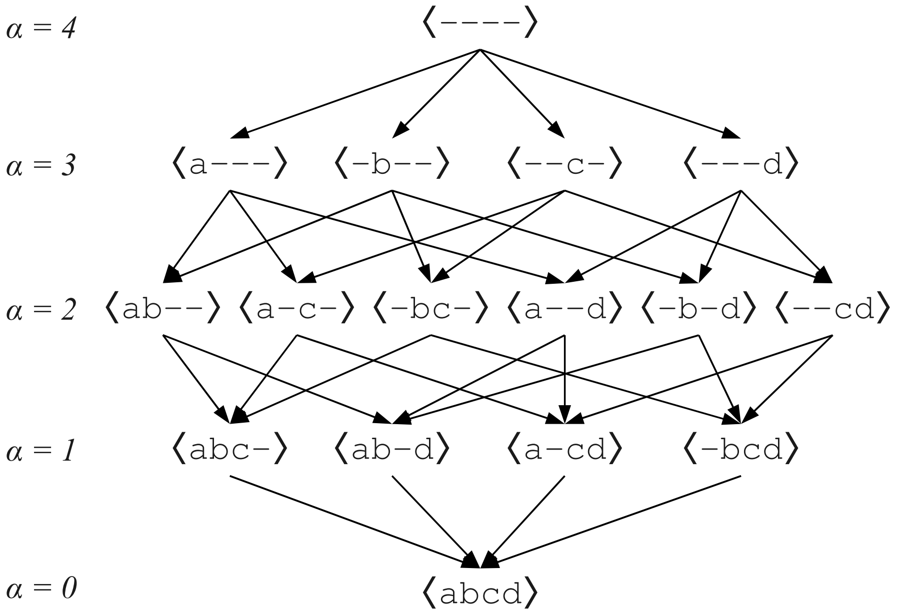


Figure 3.2: The supracontexts of ⟨abcd⟩ classified according to their abstraction α. Arrows indicate inheritance.

symbolized with the letter α in this thesis. A supracontext ⟨a--d⟩ thus has an $\alpha = 2$, as there are two values left unspecified. Given a given context of n features, the α of a supracontext is always between 0 and n, with 0 being minimally abstract and n being maximally abstract. Figure 3.2 gives all supracontexts of a given context ⟨abcd⟩ and their α-values.

Given a hypothesis such as ‘For adequate linguistic categorization behavior on phenomenon X, only relatively concrete linguistic knowledge is necessary’, we assess the predictive power of the model with the exclusion of supracontexts with a high α. If we have a set of exemplars coded for four features, we can specify that only supracontexts with an $\alpha \leq 2$ are allowed in the supracontextual analysis, so that supracontexts like ⟨--c-⟩ ($\alpha = 3$) are a priori excluded from the supracontextual analysis. The notation I will use to refer to an allowed range R_α of supracontexts is by specifying the lowest value of α allowed and the highest value of α allowed. $\alpha \leq 2$ thus means that the allowed range of supracontexts is $R_\alpha = [0 \ldots 2]$, whereas $1 \leq \alpha \leq 3$ will be marked as $R_\alpha = [1 \ldots 3]$.

Predicting the outcome of a new item without the most abstract supracontexts thus is a method of investigating the possibility of adequate linguistic behavior by an agent

		exemplars per outcome			agreements per outcome		
supracontext	α	etje	tje	pje	etje	tje	pje
⟨-Am⟩	1	0	0	2	0	0	2
⟨b-m⟩	1	1	0	1	1	0	1

Table 3.7: Homogeneous non-empty supracontexts of the given context ⟨bAm⟩ and their outcomes.

that does not have this abstract knowledge, and reflects the question whether abstract knowledge is necessary in producing and interpreting items of the category under investigation. Starting from the opposite direction, we can also operationalize research questions concerning the amount of specific, idiosyncratic knowledge necessary to produce and interpret linguistic material adequately. Does detailed knowledge help agents categorize or does it only bring about noise in the cognitive system?

Recall our four feature artificial exemplars. We can now test how well the model performs without specific knowledge by excluding concrete supracontexts with $\alpha \leq 1$, for instance. Supracontexts such as ⟨ab-d⟩, with an $\alpha = 1$, are then a priori left out of the supracontextual analysis. And similarly to assessing the same test items with decreasing degrees of abstraction allowed, we can execute experiments in which the allowed specificity is decreased with each run. Finally, combinations of restrictions on the abstraction and on the idiosyncracy are possible as well. For our four-feature exemplar, we can specify that only supracontexts with $R_\alpha = [1 \dots 3]$ are allowed, effectively excluding supracontexts with $\alpha = 0$ and $\alpha = 4$.

Our example may help clarify this approach. What is the amount of abstraction necessary to perform well on categorizating the diminutive allomorph for *baam*? Suppose we say that the maximum number of unspecified values of any supracontext is 1, or $R_\alpha = [0 \dots 1]$. For our homogeneous supracontexts, this means that no supracontexts are excluded, as both have a level of abstraction $\alpha = 1$. The prediction thus remains the same.

We can also predict the outcome using only abstract supracontexts, say of $R_\alpha = [2 \dots 3]$. This means both homogeneous supracontexts are excluded from the analysis. This, in turn, changes the prediction drastically, as none of the outcome labels has any votes and can thus be predicted any longer. The prediction then becomes a tie.

3.1.3 Evaluation

To assess how much abstract and specific knowledge is necessary to generalize towards new items, we can evaluate the performance of the model on a set of test items. If the model performs poorly given a certain range of allowed α values and it performs better on another range, we have evidence that the set of supracontexts in the latter range forms a more accurate model of a human being's knowledge than those in the former range. Thus by leaving out certain pieces of knowledge, we can test whether the model's prediction fits the actual language behavior. This fit with the way humans do it is central to this thesis. In this thesis, I evaluate the model's performance given certain settings for the range of allowed abstraction R_α by letting the model perform leaving-one-out cross-validation on a dataset. This procedure is executed for all possible settings of R_α, so that we can compare the accuracy scores of the different settings in a consistent way.

Leaving-one-out cross-validation is a method in which each of the items of a dataset is tested with all other items in the data set constituting the training set (Ney, Martin & Wessel 1997). In the case of our diminutive example, this would mean that the first case, *raam*, would be predicted on the basis of the other eighteen cases, after which the category of the second stem, *boom*, would be predicted on the basis of the first and third til nineteenth diminutives, and so forth. In total, all nineteen of the items are each predicted on the basis of the other eighteen items.

As we, the experimentors, know in advance what the category labels of the items are, we can evaluate whether the model predicts the observed category labels correctly. The crudest evaluation metric amounts to summing up the number of correctly predicted outcomes and dividing this by the total number of items. This metric is called the *accuracy score*. Applying leave-one-out cross-validation to our set of six items with no restrictions on the range of abstraction amounts to the predictions in table 3.8 in the third column. At the bottom, we can find the accuracy score for $R_\alpha = [0 \ldots 3]$, which is 0.833.

As we are interested in the predictive power of the model given different settings of the range of allowed supracontexts, we can narrow down this range. For exemplars with three features, ten ranges R_α are possible, and we will deal with the entire set of ranges quickly, but for now, the fourth and fifth column of table 3.8 give the predictions of the model given the ranges $R_\alpha = [2 \ldots 3]$ and $R_\alpha = [0 \ldots 1]$ respectively. The accuracy scores for these two models, as well as for the all-encompassing $R_\alpha = [0 \ldots 3]$ are also given at the bottom of table 3.8. What we can see here is that the highly concrete model ($R_\alpha = [0 \ldots 1]$) performs only slightly worse than the all-emcompassing model, whereas the highly abstract model ($R_\alpha = [2 \ldots 3]$) makes nine errors in total, if we take ties to be errors too.

The relative number of correct predictions often does not reflect the actual performance. As it is an overall measure, it might disguise a terrible predictive power on a low-frequency category outcome. It furthermore does not specify for which category the predictions go wrong. Perhaps both `etje` and `tje` are predicted perfectly, but the model fails to predict any of the instances of `pje` right. If we want to know where abstract knowledge is necessary and where it can be left out, we must know for which category labels the model deteriorates on different settings of R_α. Other values than the accuracy provide us with more insight in the predictive power of the model. The precision and recall, for instance, are measures based not only on the amount of

		predicted given range of abstraction R_α		
exemplar	observed	$R_\alpha = [0 \ldots 3]$	$R_\alpha = [2 \ldots 3]$	$R_\alpha = [0 \ldots 1]$
⟨rAm⟩	pje	TIE	etje	pje
⟨bOm⟩	pje	etje	etje	TIE
⟨zOm⟩	pje	pje	TIE	pje
⟨nAm⟩	pje	pje	TIE	pje
⟨wal⟩	etje	etje	etje	etje
⟨bar⟩	etje	etje	etje	etje
⟨kom⟩	etje	etje	etje	etje
⟨lam⟩	etje	pje	TIE	pje
⟨bal⟩	etje	etje	etje	etje
⟨rol⟩	etje	etje	etje	etje
⟨hol⟩	etje	etje	etje	etje
⟨bom⟩	etje	etje	etje	TIE
⟨hAl⟩	tje	etje	etje	TIE
⟨lAn⟩	tje	tje	tje	tje
⟨kOr⟩	tje	tje	TIE	tje
⟨kOl⟩	tje	etje	etje	tje
⟨lOn⟩	tje	tje	tje	tje
⟨tAl⟩	tje	tje	TIE	tje
⟨pAr⟩	tje	tje	tje	TIE
accuracy		$\frac{14}{19} = 0.737$	$\frac{10}{19} = 0.526$	$\frac{13}{19} = 0.684$

Table 3.8: Predictions for the dataset of nineteen diminutives using the leave-one-out cross-validation method given three different settings for the range of abstraction R_α.

correctly predicted good outcomes (so-called *true positives*), but also on the amount of wrongfully predicted hits (*false positives*) in the precision and on the amount of wrongfully predicted misses (*false negatives*) in the recall. The precision and recall are calculated per category label.

The *precision* is calculated by dividing the true positives by the sum of the true positives and the false positives, and thereby calculates the proportion of predicted hits of a category that are actual hits. This score reflects to what extent the model does not overgeneralize. The precision becomes lower the more false positives are made, and false positives are made if the model predicts that category label falsely, and thus overgeneralizes its knowledge. As it reflects a proportion, the precision is always between 0 and 1, and the higher it is, the better the model. For the category of pje at $R_\alpha = [0 \ldots 3]$, there are two true positives: observed cases of pje predicted to be a case of `pje`. There furthermore is one observed case of `etje` predicted to be a case of `pje`, which amounts to one false positive. Such cases are overgeneralizations: the model extrapolates too far from its knowledge about the category of `pje` to include this case, which does not belong to the category. The precision of the category of `pje` is therefore $\frac{2}{2+1}$, or 0.667.

The recall is calculated by dividing the true positives by the sum of the true positives and the false negatives, and thereby calculates the proportion of actual cases of a category that the model predicts as such. This score reflects to what extent the model does not undergeneralize. The recall becomes lower the more false negatives are made, and false negatives are made if the model fails to predict that category label if it should have, and thus fails to generalize towards that test item. Again, the recall is a proportion and therefore between 0 and 1. The higher the recall, the better the model. For the category of `etje` at $R_\alpha = [0 \ldots 3]$, there are seven true positives: observed cases of `etje` predicted to be cases of `etje`. There is also one case in which an observed case of `etje` is falsely predicted to be a case of `pje`, amounting to one false negative. This is an undergeneralization: the model fails to extend its knowledge about the category of `etje` to this specific item and thus does not capture an 'extension' that humans do make. The recall of the category of stools is thus $\frac{7}{7+1}$, or 0.875.

Typically, the harmonic mean (F-score) of the precision and recall is used as a balanced version of a global accuracy measure for a category label. It summarizes the precision and recall and its score is 'punished' if the precision and recall are far apart. All formulae are given below:

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \tag{3.1}$$

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \tag{3.2}$$

$$F = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{3.3}$$

With these accuracy measures, we can now give a more complete picture of the performance of AM on our dataset given different settings of the range of allowed abstraction. Table 3.9 gives the precision, recall and F-score per category label for the three settings of R_α, and table 3.10 gives the F-scores for all ten settings of R_α. A visualization can be found in figure 3.3, where the precision, recall and F-score for two sets of ranges are plotted per category.

The scores in table 3.9 can serve as strong indicators of what happens to the categorization process. With only abstract supracontexts allowed ($R_\alpha = [2 \ldots 3]$), the category of `pje` disappears completely: not a single case is predicted correctly. The undergeneralization of `pje` thus is complete. All these cases, now, are predicted to be formed with the allomorph `etje`, which thereby overgeneralizes strongly, as we can see by its precision (0.636). For `tje`, there is undergeneralization. All cases predicted

measure	$R_\alpha = [0\ldots3]$	$R_\alpha = [2\ldots3]$	$R_\alpha = [0\ldots1]$
precision(`pje`)	$\frac{2}{3} = 0.667$	$\frac{0}{1} = 0.000$	$\frac{3}{3} = 1.000$
recall(`pje`)	$\frac{2}{4} = 0.500$	$\frac{0}{4} = 0.000$	$\frac{3}{4} = 0.750$
F-score(`pje`)	0.571	0.000	0.857
precision(`etje`)	$\frac{7}{10} = 0.700$	$\frac{7}{11} = 0.636$	$\frac{5}{5} = 1.000$
recall(`etje`)	$\frac{7}{8} = 0.875$	$\frac{7}{8} = 0.875$	$\frac{5}{8} = 0.675$
F-score(`etje`)	0.778	0.737	0.769
precision(`tje`)	$\frac{5}{5} = 1.000$	$\frac{3}{3} = 1.000$	$\frac{5}{5} = 1.000$
recall(`tje`)	$\frac{5}{7} = 0.714$	$\frac{3}{7} = 0.429$	$\frac{5}{7} = 0.714$
F-score(`tje`)	0.833	0.600	0.833
accuracy	$\frac{12}{19} = 0.737$	$\frac{10}{19} = 0.526$	$\frac{13}{19} = 0.684$

Table 3.9: Accuracy measures for leave-one-out cross-validation categorization task on the basis of the dataset of nineteen diminutives, given three different settings of the range R_α.

to have the allomorphic category `tje` do have it, but not all observed cases of `tje` are predicted as such.

A highly concrete model ($R_\alpha = [0\ldots1]$) is better capable of predicting cases of the category of `pje`. It does so by excluding the overgeneralization of relatively abstract supracontexts predicting `etje`. We can see that this is the case by looking at the score reflecting the overgeneralization of that category, viz. its precision, which is 1.000. There thus is no overextension of `etje`. For `tje`, finally, the restriction on the level of abstraction has no effects, as compared to $R_\alpha = [0\ldots3]$. What this exercise shows us, is that the category of `pje`-allomorphs depends on very specific schemas. This fits our general idea, namely that if we are supposed to come up with an explanatory rule for this category (e.g. for second language learners), it would be that `pje` is typically produced if the nucleus is a tense vowel and the coda is an /m/. What this exercise also shows us, is that the extraction of such rules is not necessary to perform adequately on the basis of analogy with stored exemplars only.

Looking at the F-scores in all ten possible settings for R_α (table 3.10), we can see that there are four kinds of ranges. Firstly, the extreme ranges, $R_\alpha = [0\ldots0]$ (top-left) and $R_\alpha = [3\ldots3]$ (bottom-right) both predict no case correctly, and therefore have a score of $F = 0.000$ on all categories. This is typical for the model, as we shall see in the subsequent experiments. Models without any abstraction ($R_\alpha = [0\ldots0]$) cannot generalize beyond direct hits, so that only those cases that are seen before can be retrieved from memory. Only allowing for maximum abstraction ($R_\alpha = [3\ldots3]$),

		Maximum α			
		0	1	2	3
Minimum α	0	0.000 0.000 0.000	0.857 0.769 0.833	0.571 0.778 0.833	0.571 0.778 0.833
	1		0.857 0.769 0.833	0.571 0.778 0.833	0.571 0.778 0.833
	2			0.000 0.737 0.600	0.000 0.737 0.600
	3				0.000 0.000 0.000

Table 3.10: F-scores for a leave-one-out experiment on the diminutive dataset, given all possible ranges of α. The upper number in each row is the F-score of the category `pje`, the second one the F-score of the `etje`-diminutives and the bottom one the F-score for the category of `tje`.

causes only the supracontext $\langle\texttt{---}\rangle$ ($\alpha = 3$) to be available. As this supracontext is always heterogeneous, the model will have no schemas to predict the outcome of any new item on.

Between these poles, there is variation. Models disallowing abstraction beyond $\alpha = 1$, i.e. models allowing only supracontexts with zero or one features left unspecified, perform the best. Both the models at $R_\alpha = [1 \ldots 1]$ and at $R_\alpha = [1 \ldots 2]$ yield an F-score of $F = 0.857$ for the `pje`-pattern, which is much higher than the $F = 0.571$ of some of the models that do allow for more abstraction. For this artificial data set, it seems that allowing for more abstraction does not always improve the performance of the model, as working with schemas that have at most one open slot makes the model perform better than allowing for more open schemas.

Then there are two models that do not allow the more concrete supracontexts ($\alpha \leq 1$). These are the models $R_\alpha = [2 \ldots 2]$ and $R_\alpha = [2 \ldots 3]$. These are models allowing only supracontexts such as $\langle\texttt{--m}\rangle$, or in more verbose terms: 'the stem ends in an /m/'. The category of `pje` is no longer predicted, and is completely taken over by `etje`. Because of this, the precision of the latter drops, which subsequently lowers the F-score to $F = 0.600$. Interestingly enough, the category of `tje` seems relatively unaffected by the high level of abstraction: its F-score is only 0.041 lower than the F-scores at the levels where it is predicted best. Finally, the upper-right square, including

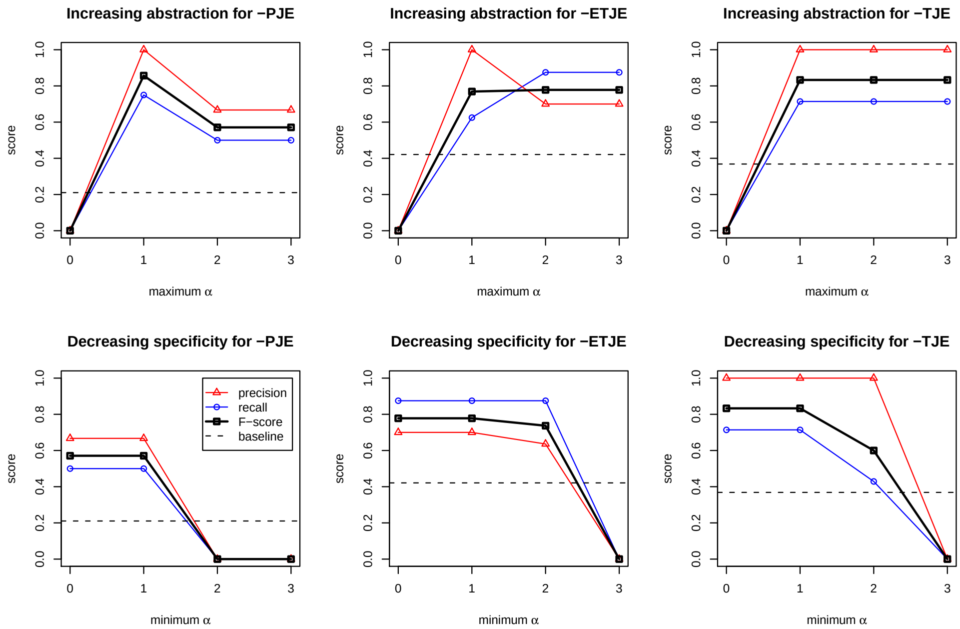


Figure 3.3: Accuracy measures for the three allomorphs on the two most salient axes, viz. increasing abstraction and decreasing specificity.

the four models $R_\alpha = [0 \ldots 2]$, $R_\alpha = [0 \ldots 3]$, $R_\alpha = [1 \ldots 2]$, and $R_\alpha = [1 \ldots 3]$, falls somewhere in between.

The two most interesting axes for the type of analysis done in this thesis, are that of increasing the allowed abstraction and decreasing the allowed specificity. In the first case, we set the minimum level of abstraction at $\alpha = 0$ and increase the maximum α stepwise. The four models corresponding to this axis are $R_\alpha = [0 \ldots 0]$, $R_\alpha = [0 \ldots 1]$, $R_\alpha = [0 \ldots 2]$, and $R_\alpha = [0 \ldots 3]$. These models form the upper row of table 3.10. The upper three panels of figure 3.3 display the effect of increasing the maximum allowed number of open slots in the supracontextual schemas. Again it is clear that the F-score for the `pje`-allomorph is predicted best at $R_\alpha = [0 \ldots 1]$. It is also visible how the other two models reach their asymptote at the same range of abstraction.

The other axis is that of decreasing the allowed specificity, and thus the allowed number of specified positions in the supracontexts. The four models corresponding to this constitute the rightmost column in table 3.10. What we see is that the categories `etje` and `tje` deteriorate relatively gracefully, but that the prediction completely fails for `pje` at $R_\alpha = [2 \ldots 3]$. This graceful deterioration is an interesting property: given only abstract supracontexts, the model can still predict many cases correctly, but many more idiosyncratic ones are lost. It is interesting to see if there are cases where the performance increases if the specificity is decreased. This would mean that a strong abstract schema is needed to predict that category, but that the prediction is distorted or made noisy if we allow for more concrete schemas that collectively dominate over the abstract schema.

3.1.4 Meeting the desiderata

Recall the five desiderata of a theory able to model the degree of abstraction in the constructicon. To my mind, AM has all the properties for fulfilling these.

A formal model AM is a formal model that generates decisions on the prediction of a category of a new item on the basis of what it has 'learned' from the exemplars in the training set.

A feature-based model AM bases the comparisons on strings of features. By means of features, phonogical, but also functional and lectal properties can be made categorial. This is a natural way for linguists to describe functions and structures. Numeric data often do not fit naturally to linguistic data. Other semantic representations could be considered and undoubtedly have advantages over features, but this would require an extension of AM which is beyond the scope of this thesis.

It is important to keep in mind that the discretization of properties into categorical values is an analyst's abstraction that may not reflect the actual representation of the meaning of a construction. Care must also be taken not to confuse the features with Jakobsonian structuralist features. In this thesis, features serve as analytic discretizations of a much more wieldy function or structure, and as such, their nature is much more instrumental to the analysis than a cognitive reality for a language user.

An unbiased model AM is not a priori biased towards the use of concrete comparisons (supracontexts with a low α), nor towards the use of abstract comparisons (supracontexts with a high α). Only the extremest levels of α on their own tend to be insufficiently open-ended (low α) or insufficiently coherent or homogeneous (high levels of α) to perform well. Given this slight bias, it is interesting to see how well the model performs on different settings in between. In the basic model ($R_\alpha = [0 \dots n]$, where n is the number of features) all play a role. By restricting the range of R_α, we can parametrize the allowed range of abstractions in the knowledge system.

A property-reduction model AM comes to a decision on the basis of the type of comparisons that strongly resemble nodes in a construction grammar network. The abstraction in the supracontexts is of the desired type, namely property reduction and not value summarization.

An unguided-abstraction model The abstract representations in AM, the supracontexts, are not created by virtue of a certain outcome type, as they are in the Varying Abstraction Model. The representation is not built up with a single purpose. From the exemplar, supracontexts specifying one value as the outcome and others specifying that value as a predicting feature can be derived. The same exemplar may thus be used for different categorization tasks.

3.1.5 Practical matters

As the available implementation of the algorithm does not allow for the parametrization I would like to execute, I implemented it myself. The programme is called `raam`, short for **r**esearching **a**bstraction in **a**nalogical **m**odeling. This programme executes the same analysis as AM, but has the option of evaluating the outcomes of the model at different settings of R_α. The programme was written in C++. A graphical user interface and a manual will soon be available on the website for this thesis.[2]

This study is a cognitive modeling study rather than a search for determinants. Because of this, I do not pay as much attention to the motivations for certain choices in the alternations as typical determinant studies do. Rather, I mostly take determinants from earlier studies. Eventually, the choice of the set of determinants is rather arbitrary, as the model should be able to perform equally well with more noise-increasing variables. A practical limitation, however, is the fact that the number of supracontexts to compare increases exponentially with the number of features. The running time of a leave-one-out analysis given one range of R_α for 11 variables and some 1000 exemplars is approximately ninety minutes on my 1Gb RAM, 1.60 GHz laptop, running a Linux operating system. The time of ninety minutes is doubled for 12 variables keeping the other factors such as number of data points equal, doubled again for 13 variables and so forth. The reasonable maximum of variables used for a data set of 1000 exemplars thus seems to be around 12 or 13. The increase in time due to an increase in the number of data points is quadratic, as for every test item in the leave-one-out analysis, all other test items have to be checked against the supracontexts of the test item. All tests were carried out using `raam` and on the computer described above. All datasets are available on the CD that accompanies this thesis, as well as on the website mentioned above.

3.2 Analogical modeling and the constructicon

So far, I have presented AM in its own terms as a categorization model that satisfies the desiderata for a model of linguistic categorization. In this section, I work out the conceptual similarities and differences between constructionist approaches and AM. Some aspects of constructionist approaches are not found in AM's ontology. I will

[2] sites.google.com/site/barendbeekhuizen/thesis

argue that they are epiphenomenal to a more basic cognitive model and that they do not belong on the analytical level of organization of the constructions used to categorize.

3.2.1 Allowing all abstractions

Different constructionist theories share the idea that the units of grammatical knowledge are form-meaning pairings, but disagree on the organization of these units. Construction Grammar only allows for non-redundant, maximally abstract storage of constructions and a complete inheritance of the properties of the mother construction in any daughter (Kay & Fillmore 1999). Usage-based approaches, such as Langacker (1987) and Goldberg (2006), display organizational networks that may contain redundant information and that allow for default inheritance, a feature whereby aspects of the mother construction can be overruled by conflicting features in the daughter construction. Redundancy is allowed as constructions often have multiple, local prototypes that are themselves used in categorizing novel items as members of that category. Frequency plays a central role in the storage of 'redundant' information (Bybee 2006).

The maximal abstraction in such a model is given by the minimal set of overlapping values of all exemplars belonging to one category, and it is not even sure that that abstraction is reached, as language users may very well operate with only schemas or constructions at a lower level of abstraction. Given a set of items only bearing family resemblances to each other, the set of overlapping properties is empty, and because empty specifications are not very useful in categorization tasks, we must assume that the cognitive agent must have some lower-level specification. Non-maximal abstraction thus is allowed by usage-based constructionist models.

Langacker's (2009) recent suggestion seems to extend this model. In his informal exposition of a usage-based model, all abstract schemas do not emerge only when they are used in comparison between two exemplars, rather they are immanent in the processing of an exemplar and are stored by the language user. Obviously, not all immanent schemas are equally useful, and some will be entrenched deeper than others. If we formalize the meaning of a construction as a set of values on certain features, we could say that the partially ordered set of this string of values is extracted in the processing, akin to the latices in figure 2.8. In figure 2.8, the items were stems and allomorphs, but we can imagine how something similar happens with the ditransitive or the cleft-construction. Next, by mapping the extracted schemas onto the existing schemas, a network, such as the one in figure 2.10 is built up and updated every time a new exemplar is processed. A conceptual advantage of this approach is that we do not have to engage in any debate about whether redundant patterns are stored as well or at what level of abstraction the highest node in the network should be. Rather, the network is composed of schemas at all levels of abstraction, and the nature of the incoming data determines what nodes or schemas will be entrenched more deeply.

AM, to my mind, is a formalism that exploits such a network by stating what nodes are useful in predicting the category label of a new item. This means that we conceptually equate nodes in the constructicon with non-empty supracontexts in AM. The entrenchment of a node is a direct function of its frequency, so that the most abstract node for a category, the one specifying no values, encompasses all members of that category. This entrenchment, however, can only be used if the pattern is sufficiently coherent. Here we touch upon the notion of homogeneity: if a string of specifications matches a new item, but shows a wide variety of different category labels associated with it, it is not a very reliable schema and it will not be used. This test can be thought of as a binary variant of cue validity or conditional probability (Beach 1974), a measure often used to determine the reliability of a property, because the homogeneity test determines the reliability or non-reliability rather than the reliability as a number between 0 and 1. In the case of the homogeneity test, we are furthermore testing the 'binary cue validity' not only of single properties, as is the case in the cue validity test, but of vectors of properties, ranging from empty vectors to fully filled ones.

In allowing initially for all abstractions, this model is different from usage-based approaches such as Goldberg (2006) and Langacker (1987), in which abstract schemas only emerge if there is sufficient evidence for them. It seems that this is a more parsimonous model, as less nodes are created and as no nodes are extracted (e.g. by children) if there is not sufficient evidence for them. However, I believe the net effect is the same, as the abstract supracontexts are only used if they are homogeneous or sufficiently coherent to be reliable predictors, such as ⟨+back⟩ with the category chair in the furniture example. Abstract supracontexts are therefore mainly useful in situations where language users can be argued to have made an abstract schema as well. The schemas of the traditional usage-based accounts thus do not correspond to the supracontexts perse, but rather to the homogeneous supracontexts.

AM resembles a true full-entry model. All supracontexts are extracted from the given context like all abstract constructions are extracted from a concrete case. These supracontexts or constructions are mapped onto the network containing the information extracted from exemplars processed earlier. According to Skousen, this network is not so much a network, but a series of comparisons made online, time and again, when processing a new item. As such, the comparisons are not part of some permanent memory organization, as I present it here. Whichever of these two implementations is used, is an empirical question: they have exactly the same outcome, and the difference lies in the fact that the stable-network version depends more on memory storage and retrieval, whereas the online-comparison version depends on online computation.

3.2.2 Links between constructions

The network that is used by AM is radically simpler than is the network used in Goldbergian construction grammar. The ontology of the model consists of schemas (or supracontexts or nodes), their complete inheritance relations and a procedure for categorizing with them. The inheritance relations can even be said to be superfluous to the categorization model, as they play no role in the categorization procedure. This set of objects and functions is strikingly simpler than the four types of links between nodes Goldberg (1995) discusses, the default inheritance procedure whereby defaults can be overruled and the reified prototypes posited as elements of the model in Langacker (1987, ch. 10) and Goldberg (2006). Obviously, all of these objects of the constructionist toolkit have theoretical significance and, more importantly, descriptive use. Nevertheless, I believe it is a mistake to confound the effects we see as linguists with the cognitive operations behind them, and that usage-based construction grammar makes this mistake by incorporating polysemy and metaphoric links, default inheritance and reified prototypes as entities in the model on par with the schemas and exemplars themselves. In the remainder of this section, I will conceptually reduce these entities to simpler interaction, thereby making the ontogeny of the model simpler, while maintaining the descriptive coverage.

First of all, let us look at the different links between constructions that are posited in Goldberg (1995, ch. 3). Apart from the 'horizontal' inheritance relations (instance links) between constructions, Goldberg also discusses polysemy and metaphoric links as well as subpart links. Subpart links are used to link two constructions of which one is a "proper subpart" of the other (Goldberg 1995, 78). Goldberg seems to mean that if one construction shares phonological, syntactic or semantic structure with another construction, the simpler one is said to be a subpart of the more complex one. As an example, she gives the Caused-Motion Construction (*I hit the cup off the table*), which is linked to the simpler Intransitive Motion Construction (*The cup fell off the table*) by a subpart link, as the two arguments of the latter are subsumed by the former (albeit in different roles).

The former two are 'vertical' links that can hold between constructions on the same level of abstraction. They capture the relation "between a particular sense of the construction and any extensions from this sense" (Goldberg 1995, 75). In a similar vein, Langacker (1987, ch. 10) discusses how upon extension, both a vertical link between the source and target of the extension as well as a horizontal link to a schema subsuming the source and target, emerge. In both Langacker (1987) and Goldberg (1995), when an extension link needs to cover too large a conceptual distance between the source and the target, for instance when the domain of experience is changed, the extension can be said to be metaphorical and a metaphorical link rather than a polysemy link is posited.

Despite the usefulness of these links in describing how constructions seem to cohere and how they have developed historically, I believe that these relations have no part in a cognitive model modeling the language users' mental processes in categorizing new items. Unless it is somehow shown that in parsing, generating or simply categorizing with constructions polysemy, subpart and metaphoric links increase the fit of the predictions with the observations, either in corpus tasks or in psycholinguistic experiments, they can be considered to be cognitive epiphenomena to a simpler underlying organization of exemplars and their schemas. While acknowledging the descriptive and analytic use of polysemy, metaphorical and subpart links, this thesis intends to prove that cognitive agents can categorize without such links as part of their cognitive system. They are not to be dismissed from the ontology of the cognitive model altogether, but we should rather place them carefully on a different level of analysis, namely either as a higher level cognitive phenomenon that is an effect of the organization and interaction of exemplars and schemas or as a linguist's attempt to draw broader generalizations about higher-level cognitive phenomena.

As Skousen (2002) acknowledges, language rules are indispensable for language education[3] and for explanatory purposes in general. They are part and parcel of our folk theory of cognition. Nevertheless, this does not mean that language users employ these rules as well. I believe the same holds for metaphoric extension links, polysemy networks and so forth. Their use lies in the fact that cognitive linguists can show how the diversity in meanings can -on an analytical level- be structured. This does not imply that language users use these mechanisms.

3.2.3 Which inheritance?

The metaphor of the network was originally motivated by the desire to display the ideas that constructions exist at different levels of abstraction, from a very general VP Construction to an instance of the *kick the bucket* idiom, and that abstract constructions are related to concrete ones by means of feature-inheritance. Because utterances are often licensed by multiple constructions, conflicts known as 'diamond problems' may arise if a construction has semantic or formal structure that mismatches the semantic or formal structure of another construction it has to unify with. Construction Grammar solves this problem by declaring the specifications of the used constructions in a vague way, so that the unification will not fail, whereas Goldberg (1995, 73-74) opts for default inheritance, the process whereby specifications of more abstract constructions are overruled if more specific constructions conflict with them.

An example is the category `bird`. One default property stored in the abstract schema is `canFly`. Say that this schema is something like ⟨ `canFly, laysEggs, hasTwoLegs` ⟩. To certain birds, however, this schema does not apply (such as penguins and ostriches, birds that have become typical through their atypicality). Default inheritance now states that at an abstract level, the default feature is present, but this feature is overruled by more concrete information about these birds, namely ⟨`cannotFly`⟩. This complicates the cognitive mechanism necessary to categorize with these constructions, as for all cases the model will have to make another check whether the default values are not overruled. Another problem is what happens if a language user tries to unify two constructions of equal abstraction, say ⟨`11-`⟩ and ⟨`-22`⟩. Here, the middle features mismatch, and unification would fail, unless one value is overruled by the other. We can, however, not decide which one is more abstract, as both are, given the formalization, equally abstract ($\alpha = 1$).

My take on this would be to leave out the concept of default inheritance from the cognitive model. Because the network necessary to categorize with Analogical Modeling in principle has a fully filled model, the node that is used in categorizing

[3]However, I believe learning the hard rules of a second language can only provide a learner with the bare basics; the idiosyncratic facts will have to be learned through examples.

the ostriches and penguins is minimally more abstract, for instance ⟨`laysEggs`, `hasTwoLegs`⟩. This means that the cognitive agent cannot employ the 'prototypical' schema, but categorization nevertheless succeeds as a somewhat more abstract specification does match the present cases. When applied to AM, the supracontexts containing ⟨`canFly`⟩ cannot be used in categorizing ostriches and penguins, so that its score will probably be lower than that for robins and pigeons, birds that can use supracontexts containing that specification. This lower score will then reflect the atypicality of the birds, so that it is in accordance with the typicality ratings. Mutatis mutandis, the same applies to the ostriches and penguins among the grammatical constructions.

I suspect the reason Goldberg feels the necessity of default inheritance, is that categorization proceeds on the basis of abstracted prototypical schemas. As with Langacker (1987), Goldberg's work suggests that reified prototypes play a major role in the categorization process. The present model reduces these prototypes to emergent constellations of nodes interacting with a test for its predictive value, as we have seen. Using the entire network in the categorization process implies that nodes at slightly different levels of abstraction exist, and that these nodes have an approximately equal number of exemplars supporting them. Because of this, a cognitive agent does not have to surpress values of one construction in combining it with other constructions if they conflict. Rather, he uses a slightly more abstract version of that construction in which these conflicting values are simply left unspecified. The argument that an extension of the cognitively simple mechanism of full inheritance, viz. default inheritance, is necessary for covering the data, loses its validity if we allow for such a full-entry network.

3.2.4 Prototypes as effects

Another feature in which the proposed network view of constructional organization differs from all strands of construction grammar is the fact that the units used to categorize are not determined a priori. Goldbergian construction grammar, as well as Langacker's Cognitive Grammar use single or multiple prototypes to represent a category. These are, as Langacker (1987, ch. 10) states, schemas, like all other schemas, but with a special status, namely that they are used in categorization. Without explicitly excluding all other nodes from the categorization process, Langacker does suggest such a thing by claiming that it would be computationally too heavy to use the entirety of a network in the categorization process. The fact that Goldberg describes prototypes as things that can be encountered (Goldberg 2006, 87) and as particular senses in a polysemy network that can be extended (p. 169-170), suggests that she implicitly assumes such a representation as well.

Langacker's (2009) recent suggestions, namely that all schemas are immanent and lie dormant in cognition until they are used in a categorization task, point in another direction. In such a view, there is no prototype as a member of the ontology of the model, or as a privilege a node or schema can have, rather (in AM terms) prototypicality is an effect that emerges from the distribution of the exemplars over the supracontexts and the usability (homogeneity) of these supracontexts. Several supracontexts have a high frequency and are homogeneous for many novel cases,[4] and thus can rightly be said to be prototypical schemas. However, this gives them no qualitatively different status than supracontexts that are used seldomly and that are only supported by few exemplars. Prototypicality thus is an emergent epiphenomenon.

Again, it is understandable how such a view could come about. Either one takes prototypes as entities in the model, a view that has been disposed with in the work of Medin & Schaffer (1978), or one regards them as analytical statements or interpretations of a researcher that are not grounded in cognitive reality, but that are used to understand the structure and mechanisms of categories and categorization or that are posited on some higher level of cognitive organization. Goldberg's characterization of prototypes seems to confound the analytical and interpretive or higher-order reality with the lower-level categorization mechanisms.

3.2.5 A dynamic grammar

In his recent exposition on a usage-based theory of language acquistion, Langacker (2009, 628) states four assumptions of Cognitive Linguistics:

- Language is seen as an integral part of cognition;
- The pivotal factor is meaning;
- Language is learned through meaningful use;
- Language is dynamic: linguistic structure is an aspect of cognitive processing.

Except for the second one, all of these apply to the Analogical Modeling approach. As I will show in section 3.2.6, AM lends itself well to models also assuming the second point. The most important point on which the assumptions of AM match the assumptions of cognitive linguistics (broadly defined) is that AM presents a general cognitive model in which categorization is fully dependent on the input data and that is built up and is updated as more exemplars are processed.

Moreover, I believe AM presents an even more dynamic perception of language than do most usage-based linguists: no prototypes are reified so that the model is maximally flexible in being continuously updated. If a category has one or multiple prototypes, we have to determine some mechanism whereby this prototype shifts if evidence is gathered that the center of gravity of a pattern has changed. In AM, prototypes are not posited, but are emergent epiphenomena that are either in the eye of the beholder or only present at some higher level of mental organization. With applications in the historical development of morphological patterns (Chapman & Skousen 2005),

[4]It is important to stress that the homogeneity of a supracontext depends on the given context and thus is possibly different for each exemplar. Translated to constructionist terms: the usefulness of a constructional schema depends on whether it is a good predictor for the current categorization task.

predictions of spelling errors (Derwing & Skousen 1994) and the correct prediction of processing times for past tense formations (Chandler & Skousen 1997), AM has shown to be a model well capable of modeling language dynamics and leakage across categorical boundaries.

3.2.6 A symbolic, multivariate grammar

AM does not present itself as a categorization algorithm for predicting form-meaning pairings. Most of its test sets are cases of allomorphy or phoneme-to-grapheme mappings, although Skousen's (1989, 97-100) sociolinguistic analysis of Arabic terms of address can be considered functional and multivariate. This functional, multivariate take is also necessary when categorizing constructions. Construction grammar takes the units of grammatical knowledge to be tuples of (at least) form and meaning. In predicting the form, we have to state which functional and other, less inclusive, formal structures motivate the constructional choice. As grammatical structures are motivated by such diverse aspects of our mental world as social relations, referential function, communicative goals and the management of the flow of information, the forms are multivariately motivated (Tomasello 2008, Grondelaers 2009).

A categorization model used for language will have to be able to deal with this multivariate nature of grammar. AM does so, by presenting us only with a non-declarative discovery procedure. The properties we use for this grammar can be whatever we think might be relevant, however remotely, for the speaker. By allowing for possibly redundant features, AM closely matches the idea that all the functions of language mentioned above may simultaneously form supracontexts used in predicting new items. Such supracontexts reflect multivariate schemas such as 'when I want to request a good friend to do something for me that involves sending me an email for which he has to do little work, I use a polar interrogative with the verb *kunnen* 'can' in the present tense'. Although we will deal with simpler supracontexts in the first two studies of this thesis, the third study, dealing with progressive auxiliaries in Dutch, actually shows how multivariate data can perfectly be used, if carefully, in modelling constructional categorization.

3.2.7 Categorization and parsing

So far, we have been discussing the use of a network in categorizing constructions as either belonging to one category or another. This categorization mechanism somehow lies at the basis of constructing and interpreting utterances using the network of constructional knowledge. However, constructing and interpreting utterances can obviously not be *equated* with one categorization task or even a series of categorization tasks. In interpreting, the sentence will have to be decomposed into its component parts before the meaning of these parts can be synthesized into the interpretation of the entire utterance. The meaning of these components, of their internal parts and of the entire utterance are interpreted using mental parsing and recombination mechanisms that go beyond mere categorization, as this task does not combine or unify meanings of parts into wholes, nor determine what the relevant parts of an utterance are.

Similarly, to produce constructions, a mental mechanism for combining meaningful chunks and patterns into sentences is necessary. Merely stating that the pattern will be a Double-Object Construction alone does not suffice to generate a sentence. The construction will have to be combined with a myriad of other constructions. At least, the parser/generator will have to be able to create wholes from meaningful parts and parts from meaningful wholes by means of some combinatorial mechanism that combines licencing constructions into a meaningful interpretation or phonological form.

Suggestions in the direction of such a model can be found in Verhagen (2009), where the notion of grammatical form is discussed, which form one conceptual basis of Beekhuizen's (2010) application of Data-Oriented Parsing (Bod 1998) in its unsupervised form (Bod 2006) to purely functional representations of utterances, effectively extracting all possible constructions from the input data and representing them in a similar total-entry network-like structure as is done here. Such a functional, usage-based parsing approach is eventually much more interesting than categorization tasks, but the development is in such an early stage of development algorithmically, that extending them to tasks on spontaneous spoken language is (yet) beyond the capabilities of most computers.

Chapter 4

Abstraction in form alternations: the Dutch dative alternation

4.1 Introduction

Being a prototypical case of a constructional alternation, the dative alternation forms a logical testing ground for the model. In this case study, I model the Dutch version of this construction. In Dutch, the two options for expressing a three-participant situation are the ditransitive pattern, as exemplified in sentence (4.1), and the monotransitive pattern with the recipient participant being expressed in a prepositional group with *aan*, as in sentence (4.2), or a number of other prepositions.[1]

(4.1) *hé dat is die vis die we jou hebben gegeven .*
hey that is that fish that we you have given .

'Hey, that's the fish we gave you.' **(fn000446.389)**

(4.2) *dus dan geef ik dat ook maar weer aan hem .*
so then give I that also but again on him .

'So then I'll give it back to him too.' **(fn000625.290)**

Another good reason to use the dative alternation as a case study, is that the determinants have been operationalized and used in categorization tasks for English, which means that we do not have to look for determinants ourselves. Because some of the principles behind these determinants are thought to be universal and because the two languages are closely related, starting with the same set of variables as Bresnan et al. (2007), seems a good starting point, but as we will see, the empirical coverage of these

[1]In examples from Dutch, the following glossing conventions will be used. 'INF' stands for infinitive, PRT stands for a particle that has no obvious translation in English, 'REDUCED' stands for a reduced form of a pronoun and 'DIM' for diminutive. The boldface code gives the file in the Corpus Gesproken Nederlands as well as the line within that file.

determinants is limited for spoken Dutch. As the Dutch alternation has been studied thoroughly as well (for a recent study, see Colleman (2006)), I will refrain from going into the debates that surround this construction, and consider it only as a categorization task.

Dative constructions such as the two above are used to express, prototypically, a transfer relation in which an agent transfers a theme to a recipient. The categorization task at hand consists of producing the recient participant as an NP (a true 'ditransitive' pattern) or as a PP (a prepositional dative). The category labels for these categories in the remainder of this chapter will be `NP` and `PP`, respectively.

Now, in categorizing cases like (4.1-4.2), what is the constructional knowledge used by speakers of Dutch? Do they use the separate associations between, for instance, the animacy of the theme argument and the category label, or do they use more specific strings of values, such as 'the theme argument is an inanimate, given personal pronoun and the verb is *ask*'. We investigate this by letting AM classify a dataset of `NP`-ditransitives and `PP`-ditransitives at different ranges of allowed supracontextual abstraction (R_α). When low levels are excluded, bigger and more specific strings of values cannot be used in the categorization, whereas the exclusion of abstract supracontexts disables the predictive use of schemas with many open slots.

4.2 The constructions and their determinants

4.2.1 The dative alternation in Dutch

The data set in this experiment consists of all `NP` and `PP`-constructions in the Netherlandic Dutch section of the syntactically annotated part of the *Corpus Gesproken Nederlands* ('Corpus of Spoken Dutch', henceforth *CGN*). This collection of data contains utterances from all parts of the corpus, except for component *n*, which contains read-aloud written language, and contains $668,260$ words. As such, the language in this data set is spoken and more or less spontaneous (ranging from spontaneous interaction to television news bulletins).

The reason for the use of the syntactically annotated part of the *CGN* over, for instance, the more homogenous section *a*, which contains only spontaneous dialogue, is that in order to extract the double-object patterns, linear search queries are not powerfull enough and generate an enormous amount of noise. By looking for all sentences in the syntactically annotated data set that contain an indirect object (`secondary edge = obj2`), a set of 2068 sentences was extracted.

From these, only the active, transitive sentences containing an indirect object expressed with a noun phrase or with a prepositional object with *aan* 'on' were kept, amounting to a total of 639 sentences, 119 of which displayed the *aan*-construction and 520 of which were cases of the double-object pattern. The prepositional dative with *voor* 'for', despite being interesting, was discarded from the data set due to its low frequency (13 cases were found among the active, transitive clauses). The preposition *tegen* 'against', co-occurring with verbs of communications such as *zeggen* 'say', was excluded as well due to its low frequency ($n = 3$). This low frequency seems somewhat artificial, and I expect that some cases of *tegen* that could be considered 'dative' in a communicative sense were not marked as `secondary edge = obj2` in the annotation. Finally, the pattern with the preposition *van* 'of', used in privative contexts (see Delorge 2009), was not marked as `secondary edge = obj2` in the syntactic annotation, and hence no cases of this pattern were extracted.

In a previous study on the syntactically annotated CGN-data (Van der Beek 2004), different figures and relative frequencies were obtained. One reason for this is that Van der Beek excluded many cases which I include, such as ditransitives with clausal theme arguments and ditransitives with fronted theme arguments. Three-participant relations expressed with other prepositions than *aan* were excluded, as their frequency in the corpus was too low to make them usable for our analysis.

4.2.2 Determinants and corpus data

The data were coded for an array of variables known to be determinants of the alternation from previous studies, such as Bresnan et al. (2007) and Colleman (2006). The following variables were used.

Abstract sense The first determinant is the abstract sense of the construction. The factor will be called `sense` in the analysis. The broad field of the 'transfer of possession' meaning associated with the ditransitive (`NP`) pattern, can be split up in several modalizations of the transfer (caused transfer, refused transfer, future transfer and enabled transfer), and extensions into communicative and other metaphorical senses (Goldberg 1995, ch. 5). Unlike more fine-grained conceptualization patterns, this schema of abstract senses is relatively straightforwardly codable in corpus data. I followed Bresnan and colleagues' coding schema, which contains the five levels `abstract`, `communication`, `transfer` (actual transfer), `future transfer` and `prevention of possession` (refused transfer). Sentences 4.1 and 4.2 above display cases of `transfer`, whereas the sentences 4.3 and 4.4 exemplify the `future transfer` and `abstract` sense respectively.

(4.3) *had je haar scherven beloofd ?*
did you her potsherds promised ?

'Did you promise her potsherds? **(fn000741.293)**

(4.4) *God wie die ook is en die toch op de één of andere manier richting geeft aan het leven ...*
God who that also is and that PRT on the one or other way direction gives to the life ...

'God, whoever He may be, who gives direction to life in one way or the other ...' **(fn000271.59)**

	sense					
	abstract	communication	future transfer	prevention of possession	transfer	**total**
NP	127	286	19	27	61	520
PP	15	68	1	2	33	119
total	142	354	20	29	94	639

Table 4.1: Frequency of the dative constructions per semantic type, as given by Bresnan et al. 2007.

	clause					
	imperative	infinitive	main	question	subordinate	**total**
NP	18	35	330	27	110	520
PP	5	12	68	9	25	119
total	23	47	398	36	135	639

Table 4.2: Frequency of the dative constructions per clause type.

As we can see in table 4.1, the outcomes are not distributed equally over the different senses. The abstract sense has a clear preference for NP datives, whereas the PP is relatively the strongest for the transfer sense. The prevention of possession-class is somewhat artificially low, as some verbs in it take other prepositions than *aan*.

Clause type This variable was not found in earlier studies, but in the exploration of the data, I found that the prepositional construction occurred more often in imperatives, questions (both wh-questions and polar interrogatives) and infinitive clauses (both main and subordinate) than in regular main and subordinate clauses. I do not have a direct explanation for this phenomenon. The variable will be called `clause`.

Collostruction effects The ditransitive alterantion is known to be governed partly by associations between the verb and one of the two constructions (Gries & Stefanowitsch 2004, Colleman 2006). The `verb` lemma thus forms and important source of information on the choice of one of the constructions. Table 4.3 gives the distribution for some verbs. An important point here is that some verbs that occur in the NP pattern do not alternate with the *aan*-preposition, but with other prepositions. *Zeggen* ‘say’, for instance, alternates with a PP-pattern with the preposition *tegen* ‘against’. Again, as these patterns had a frequency that was too low, they were not included. Note further that the verb-construction association is not directly used as a factor. Rather, the model

	verb				
	geven 'give'	*zeggen* 'say'	*vragen* 'ask'	*verkopen* 'sell'	*stellen* 'pose'
NP	113	45	45	4	5
PP	32	1	27	7	5
total	143	46	72	11	10

Table 4.3: Frequency of the dative construction with some of the encountered verbs.

uses the verb lemma simply as one value in its analogical reasoning.

Structure of the arguments Bresnan and colleagues also noted that there is a difference between the information value of the theme and recipient arguments in the two constructions. Very roughly speaking, the `PP`-dative provides more of a stage to the recipient, whereas the theme participant is more on-stage in the `NP`-dative, although the situation is not as symmetrical as it is presented here. One symptom of this different information value is the difference in accessibility of the referents. Accessibility is the inverse amount of cognitive effort the speaker thinks the hearer has to make in order to arrive at the correct referent of a nominal expression (Ariel 2001). Pronouns are markers of high accessibility, whereas full NPs headed by a definite article mark low accessibility. I coded the structure of the recipient and theme arguments (`structRec` and `structTheme`. This more or less subsumes levels used by Bresnan et al. (2007), such as 'pronominality' and 'accessibility'.[2] The following structures were encountered: `dropped argument`, `reduced pronoun`, `personal pronoun`, `demonstrative pronoun`, `reciprocal pronoun`, `demonstrative NP`, `definite NP`, `possessive NP`, `proper name`, `indefinite NP`, `quantified NP`, `clause` and `bare noun`. Some patterns can be found in the examples (4.5-4.6). Example (4.5) has a personal pronoun as a recipient, and a reduced pronoun as its theme and example (4.6) has a definite NP as its recipient, and a clausal theme.

(4.5) *anders kunnen ze kan ze 't ook aan hun vragen .*
otherwise can they can she it.REDUCED also on them ask.INF .

'Otherwise they can, she can ask it to them as well.' **(fn000733.118)**

[2] One weakness of the article of Bresnan and colleagues is that they do not discuss how they coded the accessibility of the arguments. To truly code for the values 'given' and 'non-given', one has to look at preceding and subsequent sentences and count for the number of mentions and distances between mentions, as Givón (1983) suggests. 'Accessibility' in Ariel's sense, on the other hand, is a formal symptom that resides in the NP itself and is thus much easier to code.

(4.6) *maar je kunt voor de zekerheid even aan uh aan de chauffeur vragen of*
but you can for the certainty PRT on uhm on the driver ask.INF if
die een seintje geeft .
that a signal.DIM gives .

'But if you want to be sure, you can ask the driver to give you a sign **(fn000259.123)**

The results for the two arguments are given in the tables 4.4 and 4.5, where only the encountered levels are shown. [3] It is clear from these tables that highly accessible recipients prefer the `NP` constructions. For the theme arguments, the distribution is less clear.

Animacy of the arguments For the same reason as for the accessibility difference, we can expect there to be differences in the degree of animacy between the arguments. In the `PP`-pattern, we expect on average more salient participants as recipients than in the `NP`. For the themes, we expect the more salient theme referents to be found among the `NP` construction. To code the animacy, I used the empathy hierarchy, as discussed in Silverstein (1976). As Colleman (2006, 551) discusses, it is likely that this hierarchy influences the topicality of a referent and hence the choice for `NP` or `PP` in Dutch as well. The effect of a binary distinction (animate vs. inanimate) for both object arguments proved a good predictor in Bresnan et al.'s (2007) study. Silverstein's animacy hierarchy includes the following levels, in this order: `speaker`, `hearer`, `human`, `animal`, `concrete object`, `abstract entity`. The two variables, empathy level of the recipient and empathy level of the theme, will be called `empRec` and `empTheme` respectively.

I added the intermediate level , `organization`, that is located between `human` and `object`, which was discussed in the coding schema of Zaenen et al. (2005). The motivation for including this level was that organizations seem to combine the agentivity of human beings (a political party can make promises, a board of directors can be sent something and so forth) with the lack of concreteness and spatio-temporal boundedness found with abstract entities such as *anger* or *kiss*. One further level was added for cases where the noun was clearly non-referential. I counted something as non-referential when it would be impossible to refer to the noun's referent in a hypothetical follow-up sentence. An example is *dat kind haalt je het bloed onder de nagels vandaan* 'that child is very annoying', in which one cannot refer to *het bloed* in a follow-up sentence.

In tables 4.6 and 4.7 the distribution of the outcomes over the empathy levels encountered is shown. There is no real global tendency to be found, except for the fact that the `recEmp` values of `speaker` and `hearer` clearly prefer the `NP` construction.

[3] Space limitations and aesthetic norms forced me to use some abbreviations in the tables. The abbreviations of NP-structure should be read as follows: dem. pronoun = 'demonstrative pronoun', red. pers. pronoun = 'reduced personal pronoun', $\emptyset$ = 'dropped argument'.

	structRec											
	bare noun	definite NP	demonstrative NP	dem. pronoun	indefinite NP	possessive NP	personal pronoun	proper name	quantified NP	reciprocal pronoun	red. pers. pronoun	**total**
NP	3	46	9	11	24	10	209	34	5	8	161	520
PP	2	29	7	0	15	10	23	27	1	0	5	119
total	5	75	16	11	39	20	232	61	6	8	166	639

Table 4.4: Frequency of the dative constructions per type of NP-structure of the recipient argument. See footnote 3 of this chapter for the explanation of the abbreviations.

	structTheme												
	bare noun	clause	definite NP	demonstrative NP	dem. pronoun	$\emptyset$	indefinite NP	possessive NP	personal pronoun	proper name	quantified NP	red. pers. pronoun	**total**
NP	46	115	51	14	37	9	89	8	77	3	50	21	520
PP	13	15	13	3	18	1	14	9	12	1	11	9	119
total	59	130	64	17	55	10	103	17	89	4	61	30	639

Table 4.5: Frequency of the dative constructions per type of NP-structure of the theme argument. See footnote 3 of this chapter for the explanation of the abbreviations.

This might be not so much an effect of empathy, but rather of the accessibility and shortness of first and second person pronouns.

Length of the arguments The final determinants are formed by lengths of the two arguments, as counted in words, a variable also used in Bresnan et al. (2007). Longer themes tend to be realized in the NP pattern, whereas longer recipients are found mostly in the PP pattern. As AM can only calculate with categorical, discrete levels, the length was discretized into the following levels: *short* (1 word), *medium* (2 words),

	empRec						
	speaker	hearer	human	object	organization	abstract	**total**
NP	109	162	211	27	6	5	520
PP	7	8	74	15	2	13	119
total	116	170	285	42	8	18	639

Table 4.6: Frequency of the dative constructions per encountered empathy level of the recipient argument.

	empTheme				
	human	object	abstract	non-referential	**total**
NP	5	61	424	30	520
PP	3	20	95	1	119
total	8	81	519	31	639

Table 4.7: Frequency of the dative constructions per the encountered empathy level of the theme argument.

long (3 to 5 words), *longer* (6 to 9 words) and *longest* (10 words and up). Tables 4.8 and 4.9 give the corpus data.

As we can see, longer themes do prefer the NP-construction (table 4.9), and longer recipients are more often found with the PP-dative.

Remember that as AM categorizes using local comparisons rather than global associations, the tables presented above are merely illustrative of the data in the corpus.

4.3 Results

The 639 data points were entered to the raam programme, using the squared or agreement statistics, and the evaluation metrics for all possible ranges of abstraction were calculated per category. Table 4.10 displays the PP's F-scores for all possible ranges of abstraction, and table 4.11 does so for the NP. The order in which the variables are reported in the supracontextual analyses is given in 4.7

(4.7) ⟨sense clause verb structRec empRec lengthRec structTheme empTheme lengthTheme⟩

It makes the explanation more intuitively clear if we focus on the two axes discussed earlier, viz. the increase of abstraction and the reduction of specificity. These

	lengthRec				
	short	medium	long	longer	**total**
NP	427	67	19	7	520
PP	48	40	17	14	119
total	475	107	36	21	639

Table 4.8: Frequency of the dative constructions per encountered length of the recipient argument.

	lengthTheme						
	∅	short	medium	long	longer	longest	**total**
NP	9	178	122	99	69	42	520
PP	1	54	35	11	14	4	119
total	10	232	157	110	83	46	639

Table 4.9: Frequency of the dative constructions per encountered length of the theme argument.

coincide with one row and one column from the tables 4.10 and 4.11. The top row, that has as its ranges $R_\alpha = [0 \ldots x]$, where $x = (0 \ldots 9)$, shows the effect of increasing the allowed number of open slots. The minimum number of unspecified supracontexts is specified and the maximum is a variable. We start with only completely specified schemas, and allow for one more open slot with every increase of x. If we allow for no open slots, the F-score for the PP is $F = 0.106$. Introducing schemas with one open position enhances the score to $F = 0.226$, and so forth. In figure 4.1, the left panel displays the effects on the F-scores of increasing the abstraction for the PP-pattern. For the sake of completeness, the precision and recall scores are also given. The left panel of figure 4.2 does so for the NP-construction.

The second dimension is that of a reduction of the allowed concreteness in supracontexts. This array of scores has the minimum α as its variable, and is thus the set $R_\alpha = [x \ldots 9]$, where $x = (0 \ldots 9)$. In the table, this is the right column. What happens on this dimension is that we allow initially for all concrete patterns, and first disallow the ones with no slots ($R_\alpha = [1 \ldots 9]$). As we can see in the right column of the table on the NP-pattern, this decrease has no harmful effect, as the F-score remains at $F = 0.903$. Then we also disallow supracontexts with one slot, then those with two slots too and so on. In figure 4.1, the right panel displays the effects of decreasing the specificity on the F-scores for the PP-pattern. For the sake of completeness, the precision and recall scores are also given. Figure 4.2 does so for the NP-construction.

		Maximum α									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Minimum α	**0**	0.106	0.226	0.351	0.422	0.421	0.409	0.368	0.370	0.370	0.370
	1		0.226	0.351	0.422	0.421	0.409	0.368	0.370	0.370	0.370
	2			0.342	0.422	0.423	0.409	0.368	0.370	0.370	0.370
	3				0.416	0.417	0.413	0.372	0.374	0.374	0.374
	4					0.424	0.389	0.353	0.355	0.355	0.355
	5						0.382	0.347	0.319	0.321	0.321
	6							0.277	0.288	0.286	0.286
	7								0.361	0.361	0.361
	8									0.178	0.178
	9										0.000

Table 4.10: F-scores for the `PP`-pattern at different ranges of abstraction.

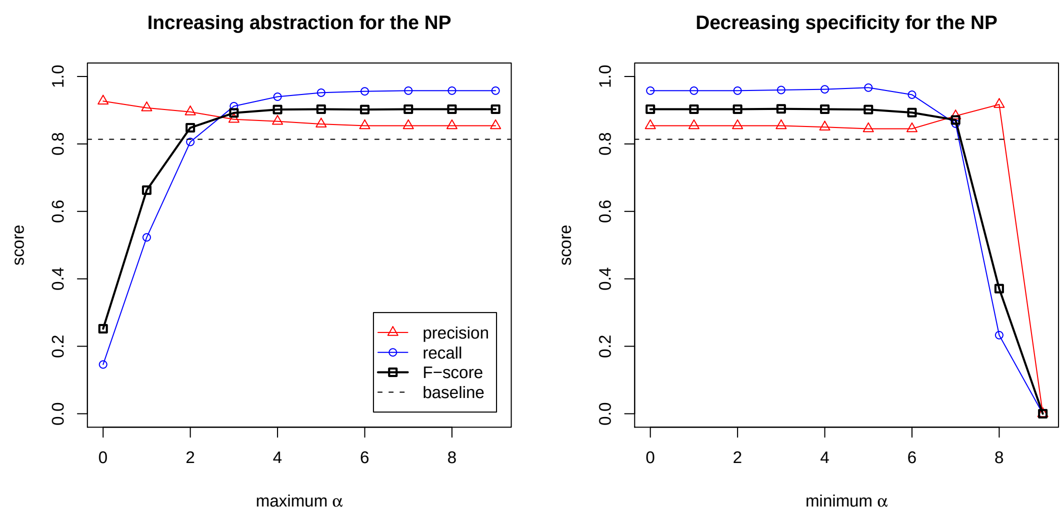


Figure 4.1: Scores for the `PP`-pattern on the dimensions of increasing abstraction and decreasing specificity.

		Maximum α									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Minimum α	**0**	0.252	0.663	0.848	0.892	0.902	0.903	0.902	0.903	0.903	0.903
	1		0.663	0.848	0.892	0.902	0.903	0.902	0.903	0.903	0.903
	2			0.848	0.892	0.902	0.903	0.902	0.903	0.903	0.903
	3				0.891	0.903	0.904	0.903	0.904	0.904	0.904
	4					0.906	0.905	0.902	0.903	0.903	0.903
	5						0.904	0.903	0.902	0.902	0.902
	6							0.894	0.893	0.893	0.893
	7								0.871	0.871	0.871
	8									0.371	0.371
	9										0.000

Table 4.11: F-scores for the NP-pattern at different ranges of abstraction.

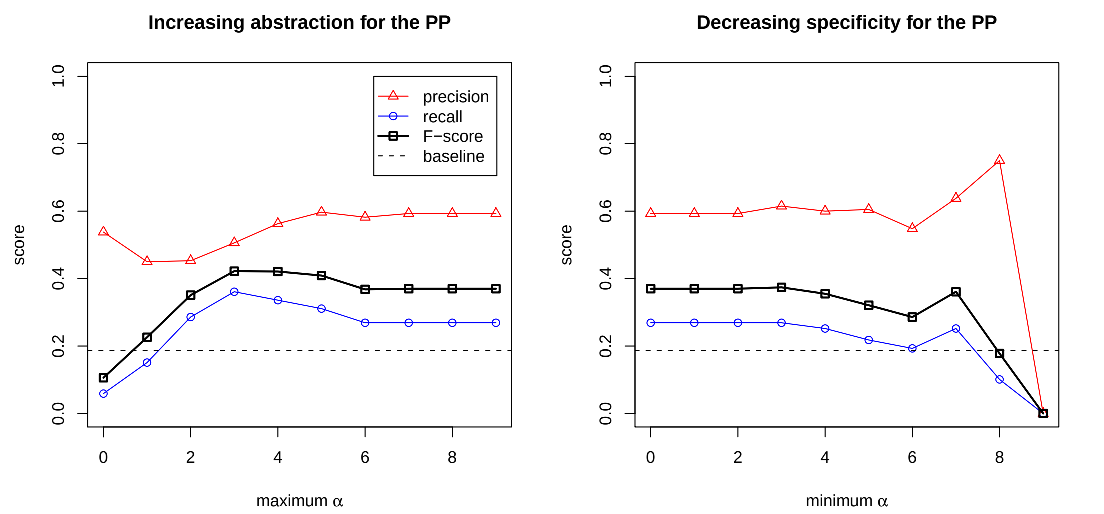


Figure 4.2: Scores for the NP-construction on the dimensions of increasing abstraction and decreasing specificity.

4.4 Discussion and qualitative exploration

4.4.1 Global interpretation

Whereas the NP is predicted nearly perfectly with F-scores around a score of $F = 0.90$, this cannot be said of the PP-pattern. The best F-scores for this construction lie around $F = 0.42$. This is nowhere near the 77% accuracy Bresnan et al. (2007) report for the PP-pattern in their spoken data set with regression analysis. A factor that might influence this score is the size of the data set. With only 119 cases of the prepositional pattern, the model may not yet be at the asymptote of its learning curve. We must keep in mind, however, that the baseline performance on the basis of blind chance is $F = 0.186$, and that optimal performance $F = 0.422$ is well over twice as high. The fact that approximately only one out of six cases is a prepositional construction, has the effect that any categorization model will have a hard time not overgeneralizing the majority category towards members of the minority category. We can see that the low scores are not due to peculiarities of AM in an analysis using Daelemans & van den Bosch's (2005) Memory-Based Learning. Such an analysis, given the same variables and the default settings of their programme `timbl`, yields highly similar scores with $F = 0.424$ for the PP and $F = 0.873$ for NP. It is clear that this set of determinants do not account for the variation between the NP and the PP-patterns in this corpus of spoken Dutch, and that more research has to be done to find out the determinants.

An indication that these variables do not cover the variation completely can already be found in the tables. Unlike the proportions reported in Bresnan and colleagues' study, the distribution of the different values over the outcomes is nowhere really outspoken enough to make a big difference. It is, for instance, not the case that all `long` recipients ($3-4$ words) are expressed in a prepositional phrase (only 17 out of 36 are). Furthermore, the PP fails to be more frequent absolutely on values on which it is relatively the strongest. The `lengthRec=long` provides a case again: relatively, 17 out of 36 cases means a preference for the PP pattern, but absolutely it does not. As the only values for which the PP-pattern is more frequent in absolute terms, are infrequent ones, such as `empRec=abstract` ($n = 18$, 13 of which display the PP-pattern), the PP-pattern does not have a firm grip on a specific subsection of the conceptual space, which it has in Bresnan et al.'s (2007) data.

Granted that we do not have the variables needed to account for this alternation, we can still make claims about the optimal degree of abstraction, albeit somewhat more carefully. In the two panels of figure 4.2, we can see the development of the three accuracy measures for the outcome NP over an increasing maximum α given a minimum of $\alpha = 0$ (left panel) and an increasing minimum α given a maximum of $\alpha = 9$ (right panel). The left panel represents ten situations in which there is one level of α more abstraction allowed with each step to the right. As we can see, the F-score starts out low, as the model cannot generalize on the basis of supracontexts that are too concrete. The precision is high, as all cases that are classified at all (i.e. that are no ties) are classified correctly. Because the range is restricted to only very concrete supracontexts, many cases do not have homogeneous supracontexts that can predict them. The recall is low as well, as most observed cases of the NP-ditransitive cannot be classified. The scores quickly rise, with the F-score reaching its peak at $R_\alpha = [0 \ldots 4]$. From the perspective of minimizing the abstraction, we could say that supracontexts in the range of $R_\alpha = [0 \ldots 4]$ are sufficient for the best possible categorization of the NP-ditransitive. Generalization and abstraction thus do not always correspond.

For the PP, the picture is very different (figure 4.1). Let us focus on the increase of abstraction first (left panel). Whereas the NP rises monotonously with the increase of abstraction, the PP displays a peak of $F = 0.422$ at $R_\alpha = [0 \ldots 3]$. After that, the F-score deteriorates slowly, settling at $F = 0.370$. The difference between these two scores is relatively big, namely 0.052. This is an interesting result, as the model that is deprived of certain (abstract) supracontexts ($R_\alpha = [0 \ldots 4]$) performs better than the model that has all possible supracontexts at its disposal. This means that more information is not always better in constructional categorization, as the abstract supracontexts in many cases seem to favour the NP category over the PP construction. This leads to the undergeneralization of the minority pattern. In contrast to the scores for the NP, the higher levels of abstraction ($\alpha > 4$) are not only not necessary, they are even counterproductive.

In the right panel of figure 4.1, we can see the effect of an increasing minimum α on the prediction of the NP-construction. Conceptually, this range corresponds to the question of the cognitive agent not storing or keeping track of concrete information or large bundles of features. Up until $R_\alpha = [6 \ldots 9]$, the scores remain stable, after which the performance deteriorates rapidly. This means that the data is sufficiently coherent to be predicted on the basis of relatively abstract knowledge. Only for extremely abstract settings ($R_\alpha = [8 \ldots 9]$ and $R_\alpha = [9 \ldots 9]$), the model is not able to predict novel cases anymore. This is due to the fact that the abstract supracontexts encompass many exemplars, and it is therefore more likely that they are heterogeneous.

A similar pattern can be observed with the PP-construction in increasing the minimum α, although the model starts deteriorating earlier. Up to a range of $R_\alpha = [3 \ldots 9]$, the model achieves reasonable scores, though never beyond a score of $F = 0.370$ for the all-supracontexts model. After that point, the performance drops, with a slight recovery at $R_\alpha = [7 \ldots 9]$. The F-score decreases almost monotonically. Less specific information thus does not lead to better predictions for the dative alternation. Combined with the results shown in the down-left panel, we can say that an optimum is reached if we allow some abstraction (maximum $\alpha = 4$). More abstraction harms the performance, as does a decrease in the specific knowledge allowed. The prepositional dative thus can only be predicted (given AM, this data set and these variables, and even then not very well so) on a relatively concrete level.

The case of the ditransitive alternation illustrates the possiblities and limitations of the model. Given a set of determinants, we may not reach outstanding scores. This makes us aware of the assumptions made when applying this technique: we rely on the data to be a reflection of the experience of a language user, we rely on the variables to reflect the actual cognitive determinants and we rely on the categorization model to reflect the way cognitive agents categorize. Nevertheless, the experiment shows interesting outcomes: the F-score of the prepositional pattern peaks at $R_\alpha = [0 \ldots 3]$. The performance on the prepositional construction furthermore deteriorates if we increase the allowed degree of abstraction, leading to the paradoxical situation that more information leads to a worse performance.

4.4.2 An example

The results presented above display a clear global tendency: for the PP-construction, storing constructions at a high level of abstraction ($\alpha > 4$) is harmful. To understand the curves in figure 4.1 better, it is insightful to analyze the predictions for a single case in more detail. The sentence in (4.8) is a case of a prepositional dative. As an exemplar, it is coded as the string in (4.9).

(4.8) *hoe legt u dat uit aan de burgers?*
how lay you that out to the citizens?

'How do you explain that to the citizens?' **(fn000177.116)**

(4.9) ⟨`communication question uitleggen def.np human medium dem.pro abstract short: PP`⟩

The exemplar in the homogeneous supracontexts of this given context that resemble the given context the most, or its 'nearest neighbor', is the sentence in (4.10), coded as the exemplar in (4.11). The only two points on which these two exemplars differ, are their main verb and the empathy level of the recipient. This item predicts the production of PP-construction, through the supracontext ⟨`communication question - def.np - medium dem.pro abstract short`⟩. Some NP-constructions also seem to be good predictors, e.g. the sentence in (4.12), coded as the exemplar in (4.13). We can see that there are multiple exemplars that are similar to the given context and that pull the new case to their outcome category.

(4.10) *kunt u die bekendmaken aan de commissie ?*
can you that known.make.INF to the committee ?

'Can you make those known to the committee? **(fn007237.45)**

(4.11) ⟨`communication question bekendmaken def.np organization medium dem.pro abstract short:PP`⟩

(4.12) *dat heb ik haar uitgelegd .*
that have I her layed.out

'I've explained that to her.' **(fn000406.253)**

(4.13) ⟨`communication main uitleggen pronoun human short dem.pro abstract short:NP`⟩

If we allow for all constructions, we can see that the model erroneously predicts this case to be a token of the NP-construction, with 275 agreements or pointers of exemplars from the NP-construction and 160 of exemplars from the PP-construction. The total amount of supracontexts is too vast to be explored in detail here, but let us zero in on what I believe to be the cause of the false prediction. There is a set of exemplars that give rise to a particularly strong cluster of supracontexts. These supracontexts are all relatively abstract, but altogether account for 228 out of the total 275 pointers predicting the representation of the 'recipient argument' NP.

The make-up of these supracontexts is as follows. As we can see in table 4.12, the verb *uitleggen* 'explain', is present in all of the supracontexts. The supracontexts furthermore display family resemblances (i.e. not all supracontexts have all values) on the following values: having the abstract semantics communication, with the theme argument being an abstract entity, that is expressed by a demonstrative pronoun and that hence is short.[4] These values form a subpart of the network that weighs heavily in the prediction. An important observation here is that this network specifies the verb and the nature of the theme argument, but leaves the specification of the recipient argument open in all cases. It is exactly this specification of the recipient that can be found with the supracontexts favouring the PP outcome. As we can see in table 4.13, most of the relatively concrete supracontexts specify features of the recipient argument, while at the same time specifying the global sense as well as features of the theme argument. When a level of abstraction is allowed in which many supracontexts have wholly undetermined recipients, we can see that the cluster of table 4.12 becomes more and more dominant.

Table 4.14 gives us the pointers per level of abstraction. Starting from a model with no abstraction and increasing the abstraction with one level per turn, we can see that there is a tie at $R_\alpha = [0 \ldots 0]$ and $R_\alpha = [0 \ldots 1]$, as there are no non-empty homogeneous supracontexts. At $R_\alpha = [0 \ldots 2]$ the PP is correctly predicted, as it is up until $R_\alpha = [0 \ldots 4]$. After that, the weight of the NP-predicting cluster becomes too great and makes the prediction switch to the NP-category. The development thus illustrates the global development we have seen in section 4.4.1: firstly, the prediction fails, then there is a level at wich the prediction is correct because of local comparisons, after which the more global supracontexts predicting the NP pattern become dominant.

4.4.3 Prototype effects and constructional organization

In the previous analysis, I showed how we can address the issue of abstraction in categorization by assessing the model at different ranges of abstraction globally, as well as on an item-specific qualitative basis. For the dative alternation, we can say that the upper bound of abstraction lies around $\alpha = 4$ given the variables used here, although

[4]It seems that some values of different features are not orthogonal to each other. Pronouns and short seem to be such cases, as well as some verbs and some global senses. It might be that this lack of independence influences the model, but it have not looked into this matter deeper.

α	supracontext	NP
4	$\langle$com - uitleggen- - -demPro abs short$\rangle$	16
5	$\langle$- - uitleggen - - - demPro abs short$\rangle$	16
5	$\langle$com - uitleggen - - - demPro - short$\rangle$	16
5	$\langle$com - uitleggen - - - demPro abs -$\rangle$	16
5	$\langle$com - uitleggen - - - - abstract short$\rangle$	25
6	$\langle$- - uitleggen - - - demPro - short$\rangle$	16
6	$\langle$- - uitleggen - - - demPro abs -$\rangle$	16
6	$\langle$com - uitleggen - - - demPro - -$\rangle$	16
6	$\langle$- - uitleggen - - - - abs short$\rangle$	25
6	$\langle$com - uitleggen - - - - - short$\rangle$	25
7	$\langle$- - uitleggen - - - demPro - -$\rangle$	16
7	$\langle$- - uitleggen - - - - - short$\rangle$	25
	total	228

Table 4.12: A cluster of supracontexts predicting the given context in (4.8). com stands for communication, demPro for demonstrative pronoun, abs for abstract. Other than this, the normal conventions apply. The numbers under NP represent the number of pointers to that outcome.

finding variables that predict the PP-pattern better may always change the performance curves in the figures 4.1 and 4.2. From this experiment, we furthermore have no evidence suggesting that reducing specificity (thus increasing the minimum allowed level of α) increases the performance. Forgetting highly specific details thus does not harm the performance, but neither does it improve the fit of the model with how the speakers in the corpus categorize. As to what the range of abstraction is at which humans reason, an educated guess would be $R_\alpha = [0 \ldots 4]$. Given this hypothesis, we can investigate whether the model displays prototype effects, and what the constructional organization might look like.

Displaying an entire network of supracontexts used to classify the ditransitive would be an impossible task, as $21,838$ different non-empty homogeneous supracontexts are used given the range of $R_\alpha = [0 \ldots 4]$. There is no way to visualize this hierarchical network, and we therefore have to focus on parts of the network that are often used. Recall that the model does not reify prototypes as entities in the model, but that prototype effects are expected to arise from the organization of exemplars. We could say that supracontextual schemas that are often applied to classify novel items can be analyzed as some sort of prototype. The clusters of most often used supracontexts can then be seen as the conceptual cores of the construction.

Figures 4.3 and 4.4 show the most frequently used supracontexts in the form of a constructional network per outcome category. Much of a true network cannot be

α	supracontext	PP	NP
3	⟨com - - defNP hum med - abs short⟩	9	9
4	⟨com - - - hum med demPro abs short⟩	4	1
4	⟨com - - - hum med demPro abs -⟩	4	1
4	⟨com - - defNP hum medium - - short⟩	9	9
4	⟨com - - - hum med demPro - short⟩	4	1
4	⟨ - - - - hum med demPro abs short⟩	4	1
$2 \ldots 4$	36 supracontexts containing information on the structure, empathy level or length of the recipient argument	30	6
	total	64	28

Table 4.13: Supracontexts within the range $R_\alpha = 0 \ldots 4$ containing information on the theme. defNP stands for definite NP, com stands for communication, hum for human, med for medium, demPro for demonstrative pronoun, abs for abstract. Other than this, the normal conventions apply.The numbers under NP and PP represent the number of pointers to that outcome.

seen, as most of the supracontexts that are most frequently used are of the abstraction levels $\alpha = 3$ and $\alpha = 4$. The inheritance relations found among the most frequent supracontexts per category are displayed with arrows in the figures. The supracontexts are displayed here as constructions, with the structural form in the upper half and the semantic and pragmatic function in the lower.

This representation clearly points out that the predictors of a formal choice do not only have to be semantic and pragmatic properties, but can also consist of structural and possibly phonological aspects. As Verhagen (2009) clarifies, signified form and signifying function on the one hand and phonological and grammatical form and semantic and pragmatic function on the other are orthogonal. A grammatical form may be a signifier of (or: have a signifying function for) another grammatical form, as can be seen from the use of structural-form entities (e.g. ‘clause type’ and ‘verb lemma’) as signifiers in predicting grammatical form alternations, such as the dative alternation.

More interesting than the inheritance relations are the clusters that emerge in the set of frequent supracontexts. For the NP-construction, we can see that many cases are predicted with schemas containing the semantic sense of communication. The transfer sense is not found among these, as for the transfer sense, there is more competition between the NP and the PP, and hence less homogeneous supracontexts. The three marked areas display three other patterns, namely that with the hearer as a recipient, the NP pattern is preferred (hatched marking), that with long themes the NP pattern is preferred (light grey marking), and that the verb *zeggen*, ‘say’, is associated the strongest with the NP-dative, alhough the verb *geven* is more than twice as frequent (dark grey marking). This is due to the fact that for *zeggen*, there is no competition

α	agreements per outcome PP	NP
2	1	0
3	21	11
4	46	33
5	53	85
6	35	104
7	4	42

Table 4.14: Pointers to the two outcomes at different levels of α for the given context in (4.9).

with the PP-pattern.[5] The global pattern is thus that the ditransitive specializes for the communicative senses, for the cases with long themes and short, preferably second person, pronouns.

For the PP-construction, we can see a rather complementary distribution. The semantic sense of choice of this pattern is the actual transfer sense (hatched marking), the recipient is expressed with a proper name or is long. Note that these two typically exclude each other as most proper names in the corpus are first names of only one word. Whether these clouds of associations actually reflect some sort of productivity, remains to be seen. It would be interesting to see how people would express situations in which they are forced to express an actual transfer with a novel verb and a recipient that somehow *has* to be expressed with a name.

Another point is that the NP-pattern is about six times as frequent in spoken Dutch as the PP pattern, and as such dominates also all kinds of sections of the conceptual space beyond the ones displayed in the figure. For the transfer sense too, the NP pattern can 'win', if only the topicality (empathy, structure, length) of the two objects is such that many supracontexts favoring that outcome apply.

We must remember, however, that these patterns are only of interpretive interest for the analyst. They have no special cognitive status in the model, as they are just some of the wide number of supracontexts that play a role in predicting the outcome. What the figures do show us, is that clouds of supracontexts that share a more abstract supracontextual schema, may be predictive without having a cluster representation to reinforce them. Suppose an agent has the schemas $\langle$ab-$\rangle$, $\langle$a-c$\rangle$, and $\langle$ad-$\rangle$. The further use of a schema $\langle$a--$\rangle$, even if it is homogeneous may sometimes turn out to be harmful, as it can lead to overgeneralization. In other cases, the absence of such an abstract schema does not lead to an improvement of the predictions on its own, but it rather seems that disallowing all supracontexts at that level of abstraction leads to a

[5]The NP-realization does in fact alternate with another pattern, viz. the prepositional pattern with *tegen* 'against'. This pattern occurred insufficiently often in the corpus data to be taken along.

more accurate prediction.

The prototypicality of a schema thus can be said to reside in its applicability. The prototypicality of an exemplar, a concrete case, lies rather in the certainty with which it can be classified as a member of the one category or the other. Given that there are 400 pointers pointing towards a verbalization of the recipient as a noun phrase argument, against 0 pointing towards a verbalization as a prepositional phrase, then the degree of certainty is much higher then when there are 250 pointing towards the former and 150 towards the latter. The case we discussed thus does not belong to the most prototypical PP-cases, as there are cases in which the PP is more strongly predicted, for instance sentence (4.14), which has at $\alpha = 0 : 4\,223$ pointers for the PP-construction, and none for the NP. In the same range of abstraction, a case like (4.15) is a non-prototypical NP-construction with 83 PP-pointers against 90 NP-pointers.

(4.14) *maar ik geef les aan uh kinderen van IVBO VBO (...)*
but I give lesson to uh children of IVBO VBO (...)

'But I teach children in the IVBO and VBO layers of eduction' **(fn000086.63)**

(4.15) *ik geef dat unithoofd uh geef ik alle alle roem .*
I give that unit.head uh give I all all honour .

'I give all credit to that unit head.' **(fn000979.160)**

Finally, cases that are wrongfully predicted, are in the light of prototypicality very atypical cases: whereas the model predicts them to belong to category B, they were observed to be members of category A. This directly points to the situation AM is known to be able to deal with. Even given that we have the set of determinants, we do not expect a perfect score, as "no language is tyranically consistent. All grammars leak." (Sapir 1921, 38). In the case of the two constructions investigated here, there is massive leakage. In many cases, PPs are predicted to be NPs, even if only the most local supracontexts are taken into account. One possible reason for this leakage could be the fact that NP-patterns occur six times as often, but we would not expect the leakage to be this big, as it would mean that the category would disapear within no time. From that fact, we can conclude that we have not found the right determinants yet, as the PP-category has survived for quite some time. Leakage is expected, but not in this order. This fact, however, does not restrict the analysis of the methodology for investigating abstraction too much, although one reservation is that if we find a determinant that strongly increases the performance, the abstraction curves of figure 4.1 and 4.2 may change drastically.

4.5 Conclusion

As a first exploration of the working of the method, modeling the ditransitive alternation has proved to be a useful exercise. Although we do not have all determinants, the present determinants suggest that too much abstraction causes the `PP` to be pushed away by the `NP`. In order for the minority pattern of the `PP` to be predicted somewhat better, broad abstractions are harmful.

The abstraction level necessary to make fairly accurate predictions is already reached pretty early on. With schemas that have four open slots (out of a possible nine), the model makes its best predictions of the language users' actual behavior. More abstraction than this is not useful for either construction and even harmful for the `PP`.

Reducing specificity makes little difference for the `NP`-pattern. Apparently, this construction is so coherent that it allows for broad abstractions over the exemplars. The `PP` starts performing worse relatively quickly. It seems that the only way a minority category can maintain itself in the conceptual space, is when the broad abstractions that often point to majority categories, are excluded.

One determinant lacking from this analysis was word order. I found it hard to decide whether or not to use word order as a predictor of the outcome. My main reason for this was that word order, especially in the kind of language I investigated (viz. spoken language) is a product of both planning and incremental production. Given that latter foundation of word order, it seems awkward to use word order as something the speaker plans ahead when categorizing the conceptual situation to be one of the two dative constructions. The givenness of the participants, which leads to different structural NP-types, and the choice of the verb seem to be more likely candidates to be planned before the speaker starts uttering the sentence.

Apart from the problem of not having the right set of determinants, another problem is that we must somehow find out whether the observed pattern is a property of the model itself, or that it resides in the data. If the former is the case, we would expect the model to display similar curves for other data sets. In the latter case, the range may vary from alternation to alternation. To investigate this, and to gain further insight in the mechanisms of AM as applied to syntactic alternations, let us look at some other cases.

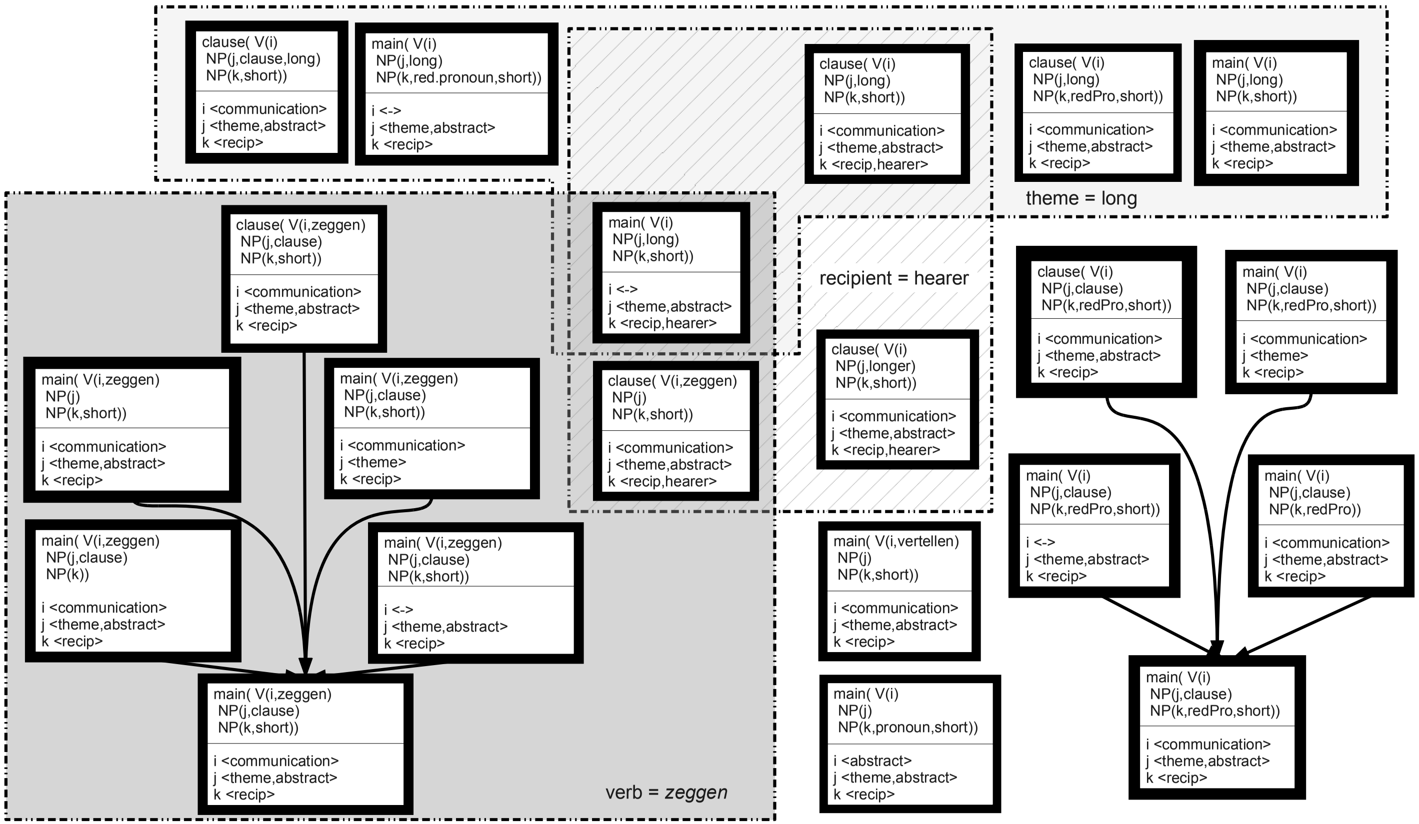


Figure 4.3: The most frequently applied supracontexts in the prediction of NP-ditransitives. The thickness of the boxes corresponds with their frequency, with one point of thickness standing for a frequency of two.

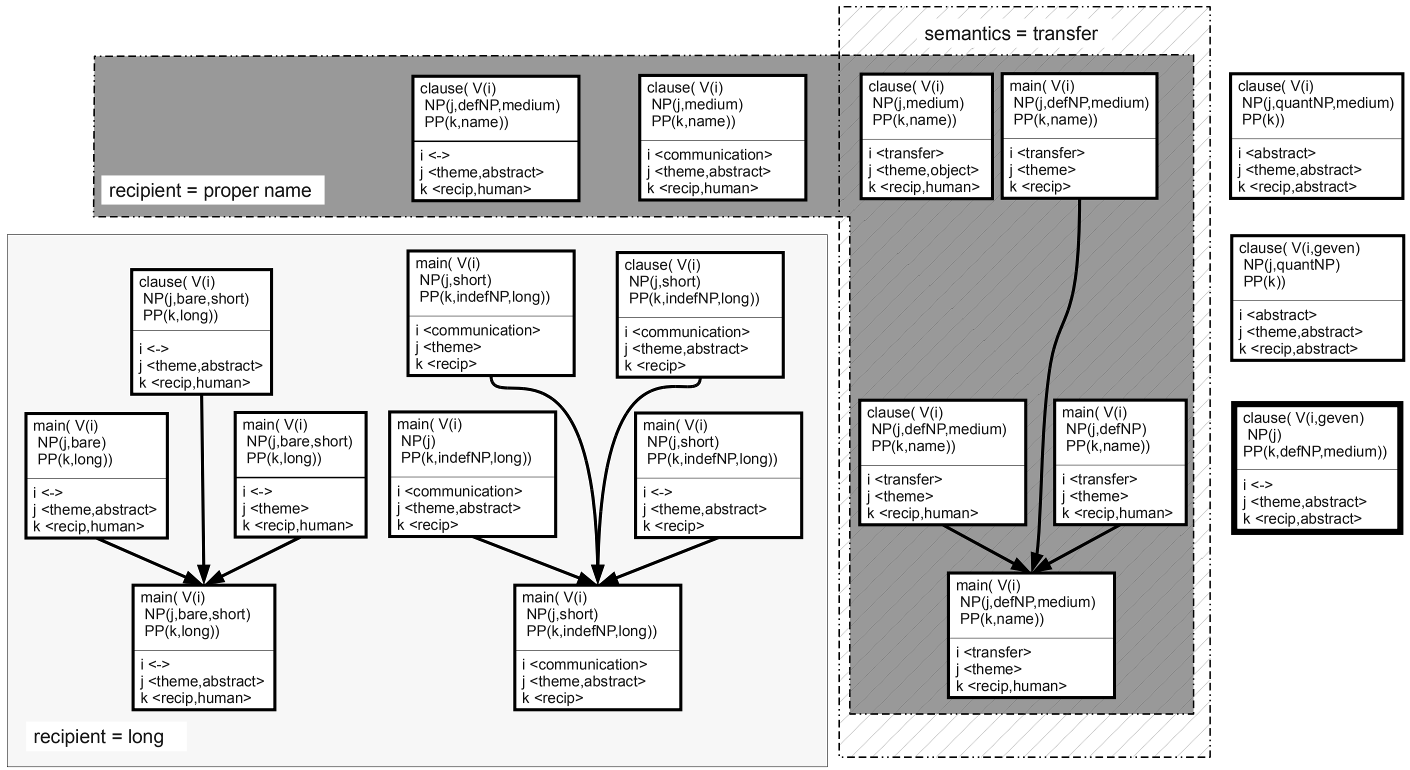


Figure 4.4: The most frequently applied supracontexts in the prediction of PP-ditransitives. The thickness of the boxes corresponds with their frequency, with one point of thickness standing for a frequency of one.

Chapter 5

Abstraction in disambiguation tasks: relative clauses in Dutch

5.1 Introduction

Linguists interested in the function of a grammatical pattern often try to find out what the function is by looking at the difference between that construction and constructions that are functionally and formally highly similar. A prototypical case of determining the function or functions of a construction by comparison is that of the dative alternation, which was explored in the previous chapter.

A different kind of study, also involving form and function, has not gained as much attention in the constructional literature, namely that of disambiguation. Where the determinant studies predict a grammatical form from a set of functions (such as animacy and accessibility) and less inclusive forms (such as verb lemmas and the length of the constituents), in a disambiguation study, the form and the less inclusive functions (such as characterizations of the participants of an argument structure construction) are used to predict the global function of a pattern. This is important, as language users are able to analyze new sentences as well and as thus far, the issue of (potentially) homonymous constructions has not been dealt with properly within the constructionist framework. As the mental organization of constructions is used in both producing and interpreting language, language users should be able to disambiguate with them as well.

In developing a constructional account of disambiguation, the same questions raised in the previous chapters apply: do language users need to extract and employ abstract rules to interpret new items, or does the use of only fine-grained regularities suffice? And on the other side: is it necessary for a language user to remember specific, fine-grained details of the sentences or can a language user rely on broad generalizations? This case study addresses the questions of the effect of abstraction and that of concreteness for a disambiguation problem in Dutch, namely that of the grammatical role of the relative pronoun in transitive relative clauses.

5.2 The constructions and their determinants

5.2.1 Transitive relative clauses in Dutch

In Dutch, a phrase like (5.1) can mean both 'the boys that help the construction workers' and 'the boys the construction workers help', where *die* is the grammatical subject in the former interpretation and the direct object in the latter. I will call the interpretation category in which the relative pronoun is the subject the `subject` reading, whereas the interpretation in which the relative pronoun is the direct object of the relative clause will be called the `object` category.

(5.1) *de jongens die de bouwvakkers helpen*
the boys that the construction workers help

'the boys that help the construction workers'
'the boys the construction workers help'

The ambiguity of transitive relative clauses has mainly been studied from a psycholinguistic perspective, as the topic has to do with linear, incremental processing, as well as with a hierarchical interpretive structure, in which the roles of the participants differ between the two interpretations. Important work on the incremental processing of transitive relative clauses has been done by Mak and colleagues (Mak, Vonk & Schriefers 2002, Mak, Vonk & Schriefers 2006). Given the observation that `object` relative clauses are read slower (in Dutch, as well as English, French and German), and hence probably harder to process, Mak and colleagues wonder what might cause this. By systematically manipulating the animacy of the two participants, Mak and colleagues found that not so much the structural difference between subjects and direct objects causes this difference, as psycholinguists with a generative orientation assumed, but that the animacy of the antecedent participant was crucial to the reading time. In clauses with animate antecedents, the `object` clauses were indeed processed slower than the `subject` clauses, but this effect was absent if the antecedent of the relative pronoun was inanimate. Mak et al. concluded from this that object relative clauses are only processed slower if there is possible ambiguity, e.g. in cases where there are two animate participants (as in example (5.1) above).

The reading time experiments showed that it took the readers more time to make a decision on an interpretation of the clause if both participants are animate, than when one of the two participants is animate while the other is not. Most strikingly, `object` clauses with inanimate antecedents were processed equally easily as `subject` clauses with an animate argument within the relative clause. The effect can thus be attributed to animacy rather than to clause type, although a follow-up experiment in Mak, Vonk & Schriefers (2002) shows that the `subject` reading is the default in the case of two animate participants. In our experiment, the incremental processing of the clauses is not the central issue and we rather look at the application of constructional patterns in the interpretation of transitive relative clauses. In a small aside, however, I will discuss informally how the techniques of Analogical Modeling can be employed to model some of Mak, Vonk & Schriefers's (2002) results.

5.2.2 Determinants and corpus data

The data for this experiment consisted of all transitive relative clauses in component *a* of the *Corpus Gesproken Nederlands* ('Corpus of Spoken Dutch'), of which only the Netherlandic section was used. As 'transitive', I considered all cases in which two arguments of the relative clause were explicitly uttered, even if the object was semantically closer to an indirect object. The used section of the *Corpus Gesproken Nederlands* consists of $1,747,789$ words in transcriptions of spontaneous conversations. In total, 975 transitive relative clauses were found, 466 of which displayed a subject reading and 509 of which had an object reading. Only in a few cases, the interpretation was not clear on first sight, but after inspection of the context, all of the clauses could be encoded for the outcome.

Before we take a look at the determinants, I would like to introduce some terminology. The antecedent of the relative clause is called the *antecedent*, and the argument expressed within the relative clause I call the *argument*. When talking about the *arguments* in plural, I mean the antecedent and the argument together.

Number and agreement In many cases, the ambiguity is neutralized by a difference in agreement. If one of the two participants agrees in number with the finite verb, whereas the other does not, speakers can make the choice of an interpretation on a formal basis. Examples can be found in (5.2-5.3)

(5.2) *de jongen die de bouwvakkers helpen*
the boy that the construction workers help

*'the boy that helps the construction workers'
'the boy the construction workers help'

(5.3) *de jongens die de bouwvakker helpen*
the boys that the construction worker help

'the boys that helps the construction worker'
*'the boys the construction worker helps'

The first three variables we thus code, are the grammatical number of the antecedent (`nAnt`), the grammatical number of the argument (`nArg`) and the grammatical number of the finite verb in the relative clause (`nVerb`). I chose not to code agreement directly, as it can be expected to follow from these lower-level formal markers of grammatical number. Tables 5.1 and 5.2 summarize the corpus data for the variables. As expected, in cases where the finite verb does not agree in number with that participant, that participant is almost certainly not the subject. We can see this in

	nVerb		
nAnt	plural	singular	**total**
plural	67/186	164/2	138/188
singular	71/2	207/276	288/268
total	138/188	371/278	509/466

Table 5.1: Numerical agreement between the argument and the finite verb in the relative clause. The first number in each cell is the frequency of subject relative clauses displaying that combination, the second the frequency of object relative clauses with that combination.

	nVerb		
nArg	plural	singular	**total**
plural	138/40	0/40	138/80
singular	0/148	371/238	371/386
total	138/188	371/278	509/466

Table 5.2: Numerical agreement between the antecedent and the finite verb in the relative clause. The first number in each cell is the frequency of subject relative clauses displaying that combination, the second the frequency of object relative clauses with that combination.

the fact that if `nArg = plural` and `nVerb =singular`, there are no `object`-interpretations, as this would mean that the subject and the finite verb disagreed in number. Interestingly enough , there seem to be exceptions. There are four cases where antecedent does not agree with the finite verb, but is nevertheless the subject. Two of these are cases in which the antecedent is semantically plural, but not formally, as in *Nike die schoenen maken* 'Nike that shoes make', in which *Nike* refers metonymically to the employees or designers of Nike. The verb agrees here with the object *schoenen* 'shoes' only, but we nevertheless understand that the clause has a subject reading. Furthermore, in one case the verb agrees with neither the argument nor the antecedent. These are interesting cases for the model: how does it, despite lacking formal evidence or even having 'false' evidence, predict the right outcome?

Arguments A second set of determinants comes from the word form of the antecedent and the argument. These factors will be called `ant` and `arg` respectively. These factors might be of influence for two reasons. Firstly, case-maked personal pronouns can disambiguate a clause, as the nominative case can only be used to express subjects, whereas the accusative case is reserved for the rest of the function personal pronouns occur in. Sentences (5.4-5.5) give examples. Again, instead of coding this feature straight away, we let it emerge from the word forms. Secondly, certain words may be more strongly associated with one of the constructions than others. An example is *ding* 'thing', which is the most frequent antecedent in object clauses, but does not surface in the top 5 of the subject clauses.

(5.4) *de jongens die we helpen*
the boys that we help

*'the boys that helps us'
'the boys we help'

(5.5) *de jongens die ons helpen*
the boys that us help

'the boys that helps us'
*'the boys we help'

The two features were coded not as word *types* or *lemmas*, as different pronominal forms (reduced forms, case-marked forms) would be reduced to each other then, but as seperate items per encountered transcribed form in the CGN. Table 5.3 displays the most frequent five word forms of the antecedent for each of the outcomes and table 5.4 does so for the argument.

Animacy Obviously, real world knowledge forms another major determinant of the interpretation, and this is where Mak et al.'s animacy studies come in. Formally, a case like (5.6) is equally ambiguous as the case in (5.1), but our world knowledge disfavors the interpretation in which the babies take the boys hostage. In example (5.7), it is the other way around. Here, the sentence is also formally ambiguous between the two interpretations, but our world knowledge favors the interpretation in which the criminals are the agents and hence the grammatical subject of the relative clause. And by extension, the same argument applies to cases that we would only hesistantly call ambigous, such as example (5.8), in which it is clear that the boys sell the sailing ships and not the other way around. In this latter case we also have a clear difference in animacy between the two arguments.

(5.6) *de jongens die de baby's gijzelden*
the boys that the babies held.hostage

'the boys that held the babies hostage'
'the boys the babies held hostage'

(5.7) *de jongens die de criminelen gijzelden*
the boys that the criminals held.hostage

'the boys that held the criminals hostage'
'the boys the criminals held hostage'

`subject`	`object`
mens 'human' (92)	*ding* 'thing' (58)
iemand 'someone' (59)	*mens* 'human' (24)
degene 'the one' (40)	*iemand* 'someone' (9)
man 'man' (20)	*jongen* 'boy' (9)
jongen 'boy' (10)	*woord* 'word' (8)
179 different forms	326 different forms

Table 5.3: The most frequent values for the antecedent of the relative pronoun (`ant`) per interpretation type.

`subject`	`object`
dat 'that' (51)	*je* 'you' (129)
't 'it' (35)	*ik* 'I' (125)
zich 'oneself' (15)	*ze* 'they/them' (65)
het 'it' (14)	*we* 'we' (58)
dit 'this' (11)	*'k* 'I' (25)
239 different forms	44 different forms

Table 5.4: The most frequent values for the argument inside the relative clause (`arg`) per interpretation type.

(5.8) *de jongens die zeilschepen verkopen*
the boys that sailing.ships sell

'the boys that sell sailing ships'
'the boys sailing ships sell'

There is an assymmetry between the types of entities that fill the two arguments of the transitive construction (Hopper & Thompson 1980). Typically, a more powerful entity acts on a less powerful entity. In terms of the animacy of the participants, this means that the participant ranked higher on the scale of animacy is more likely to be the agent, and hence the grammatical subject. This effect can be seen in examples (5.6) and (5.8), where we could interpret the former as an object reading more easily than the latter. The animacy of the antecedent and the argument are thus taken as the fifth and sixth determinant of the outcome, and are coded using the empathy scale used in the previous chapter (Silverstein 1976). The feature labels are `empAnt` for the empathy level of the antecedent and `empArg` for the antecedent of the argument. Again, I included Zaenen et al.'s (2005) level of `organization`.

	empArg							
`empAnt`	sp	hr	hum	org	ani	obj	abs	**total**
sp	0/0	0/0	0/1	0/0	0/0	0/1	0/0	0/2
hum	52/7	18/7	15/51	0/4	0/2	0/107	1/107	87/385
org	2/0	0/0	1/0	0/5	0/0	0/7	0/23	3/35
ani	2/0	3/0	0/0	0/0	0/1	0/0	0/3	5/4
obj	77/1	67/0	43/1	1/0	0/0	1/7	0/12	189/21
abs	93/2	67/0	59/1	2/0	0/0	2/0	2/16	225/19
total	227/10	155/7	118/54	3/9	0/3	3/122	3/261	509/446

Table 5.5: Frequency of the interpretation type per combination of empathy levels of the antecedent (`empAnt`) and the argument (`empArg`). The first number in each cell gives the frequency of the `object` relative clauses displaying that combination, the second for the `subject` relative clauses. The abbreviations are to be read as follows: sp = speaker, hr = hearer, hum = human, ani = animal, org = organization, obj = object, abs = abstract. Unused levels are left out.

For the two features `empAnt` and `empArg`, the best way to explore the corpus is to see what combinations of the two occur. In table 5.5 we can see these combinations. One pattern that is immediately clear, is the animacy hierarchy. Of the two arguments, the most animate one is typically the subject. Silverstein's hierarchy runs from left to right and from top to bottom on the two axes. I placed the level of organization above that of animal, but I am not entirely confident as to whether this is the right choice. Effectively, it does not matter, firstly as the levels are coded categorically and not in some ordinal fashion. Secondly, because no cases of animals acting upon organizations or vice versa were found.

As we can see, in the vast majority of the cells that combine participants of different animacy levels, the higher-ranked participant is the grammatical subject of the clause, whether it is the antecedent or the argument. An interesting observation here is that in all of the cases of equal empathy between the two participants, the most frequent grammatical role of the antecedent is the `subject`.

Accessibility Mak, Vonk & Schriefers (2002) refer to a corpus study on German (Zubin 1979), who found that the more salient an entity is to the interactants, the more likely it is to be the subject. One aspect of salience is the empathy hierarchy we have already seen, but the information value of the two arguments may provide us with another symptom of the discourse prominence. Mak, Vonk & Schriefers (2002, 53) state that the relative pronoun is always more 'given' than the argument in the relative clause. I doubt that this is the case, as the referent of indefinite pronominal

	structArg													
	bare noun	clause	definite NP	demonstrative NP	dem. pronoun	indefinite NP	indefinite pronoun	possessive NP	personal pronoun	proper name	quantified NP	red. pers. pronoun	reflexive pronoun	**total**
`object`	0	0	4	1	8	4	4	2	182	8	1	295	0	509
`subject`	60	14	35	29	62	76	23	15	38	11	37	49	17	466
total	60	14	39	30	70	80	27	17	220	19	38	344	17	675

Table 5.6: Frequency of the interpretation types per accessibility level of the clause-internal argument.

	structAnt										
	bare noun	definite NP	demonstrative NP	dem. pronoun	indefinite NP	indefinite pronoun	possessive NP	personal pronoun	proper name	quantified NP	**total**
`object`	10	127	145	1	129	17	8	3	5	64	509
`subject`	7	49	55	0	149	116	10	4	21	55	466
total	17	176	200	1	278	133	18	7	26	119	675

Table 5.7: Frequency of the interpretation types per accessibility level of the antecedent.

antecedents such as 'someone who ...' and indefinite NPs such as 'some guy who ...', is typically non-given. It thus seems an interesting additional value to code the structure of the two noun phrases and use them as variables as well. I used the same levels as in the previous chapter, again roughly following Ariel (2001). Remember that in principle the model should be able to model the variation using also variables that do not predict very well, so that if the model does not get better, it is not expected to get worse either. Tables 5.6 and 5.7 display the results.

The results are as expected: `object` relative clauses most often have a personal pronoun, either full or reduced as the argument and a more elaborate structure as the antecedent. Examples are the full NPs headed by definite articles and demonstratives, where the `object` interpretation occurs twice as often as the `subject` interpretation. Strikingly, the `object` clauses hardly display any quantified, indefinite pronouns (such as 'someone') in their antecedent, unlike `subject` clauses. Another remarkable observation is that for indefinite NPs as antecedents, either quantified or not, the corpus data shows hardly any difference between the two relative clause types.

Verb The penultimate determinant is the `verb`. The main verb of the relative clause is included, as subtleties of the combination of animacy and word form of the two participants may display in combination with the verb. It furthermore incorporates possible conventional item-effects, something which is shown to be present with other constructions as well (Gries 2003). The main verb of each relative clause was coded in the lemma form. The most frequent ones per outcome are shown in table 5.8

Word order Finally, the word order of the relative clause, is an interesting determinant of the interpretation. As Mak, Vonk & Schriefers (2002, 67) note, the word order in the relative clause in which the argument does not immediately follow the relative pronoun cannot yield an `object` reading. Consider the sentences (5.9) and (5.10). Whereas the first one only allows for the `subject` interpretation, the latter allows for both interpretations. As this is not a thesis about determinants, the precise explanation of this phenomenon is left open here. The word order is coded as the elements that are found between the relative pronoun and the finite verb. The two elements that are coded are 'arg' for the argument and 'o' for other (adverb, PP, particle), where 'o' can contain any number of other elements. The argument followed by a PP, an adverb and three particles would thus be coded with the value 'arg+o'.

(5.9) *de jongens die gisteren de criminelen gijzelden*
the boys that yesterday the criminals held.hostage

'the boys that yesterday held the criminals hostage'
*'the boys the criminals held hostage yesterday'

(5.10) *de jongens die de criminelen gisteren gijzelden*
the boys that the criminals yesterday held.hostage

'the boys that held the criminals hostage yesterday'
'the boys the criminals held hostage yesterday'

`wordOrder` has six attested values in the corpus, as can be seen in table 5.9. The phenomenon noted earlier, namely that when anything precedes the argument in the relative clause, the relative pronoun fulfills the subject role, stands out: in all 84 cases of 'o + arg' and all 10 of the 'o + arg + o', the relative pronoun is the subject of the clause. Less clear is the case of the order 'arg + o', in which object readings occur almost twice as often. The 'simple' pattern 'arg' displays no preference for either outcome. Finally, there are two values found only with relative clauses containing complement taking predicates. In these cases the complement clause, which in my corpus was always the direct object, is always placed after the verb cluster. The Mittelfeld between the relative pronoun and the finite verb can thus be either empty, or filled by another element.

`subject`	`object`
hebben 'have, own' (87)	*hebben* 'have, own' (105)
doen 'do' (44)	*kennen* 'know (someone)' (39)
maken 'make' (17)	*doen* 'do' (38)
afluisteren 'listen' (12)	*krijgen* 'get' (23)
geven 'give', *zeggen* 'say' (11)	*zien* 'see' (19)
176 types	156 types

Table 5.8: The most frequent values for the main verb of the relative clause (`verb`) per interpretation type.

	`wordOrder`						
	arg	arg+o	o+arg	o+arg+o	o	$\emptyset$	**total**
`object`	250	259	0	0	0	0	509
`subject`	250	110	84	10	5	7	466
total	500	369	5	84	10	7	675

Table 5.9: Frequency of interpretation types per word order pattern between the relative pronoun and the finite verb of the relative clause. The abbreviations are to be read as follows: arg = NP-argument, o = other element.

5.3 Results

The dataset with the eleven variables was submitted to the `raam` programme using the squared statistics. Tables 5.10 and 5.11 give the F-score for all different ranges of abstraction, with the columns specifying the highest number of values left open for each supracontext (maximum α), and the rows the lowest number of values left open (minimum α). The figures 5.1 and 5.2 display the curves per category on the two most interesting axes, viz. that of increasing abstraction and decreasing specificity. When, from now on, exemplars are reported in this chapter, the order of the feature-values is as given in (5.11).

(5.11) ⟨`ant structAnt empAnt nAnt arg structArg empArg nArg verb nVerb wordOrder` ⟩

	Maximum α											
Minimum α	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
0	0.06	0.33	0.84	0.95	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98
1		0.33	0.84	0.95	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98
2			0.84	0.95	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98
3				0.95	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98
4					0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98
5						0.97	0.97	0.97	0.98	0.98	0.98	0.98
6							0.97	0.97	0.98	0.98	0.98	0.98
7								0.97	0.98	0.98	0.98	0.98
8									0.98	0.98	0.98	0.98
9										0.97	0.97	0.97
10											0.78	0.78
11												0.00

Table 5.10: F-scores for the `object` relative clauses at different ranges of abstraction.

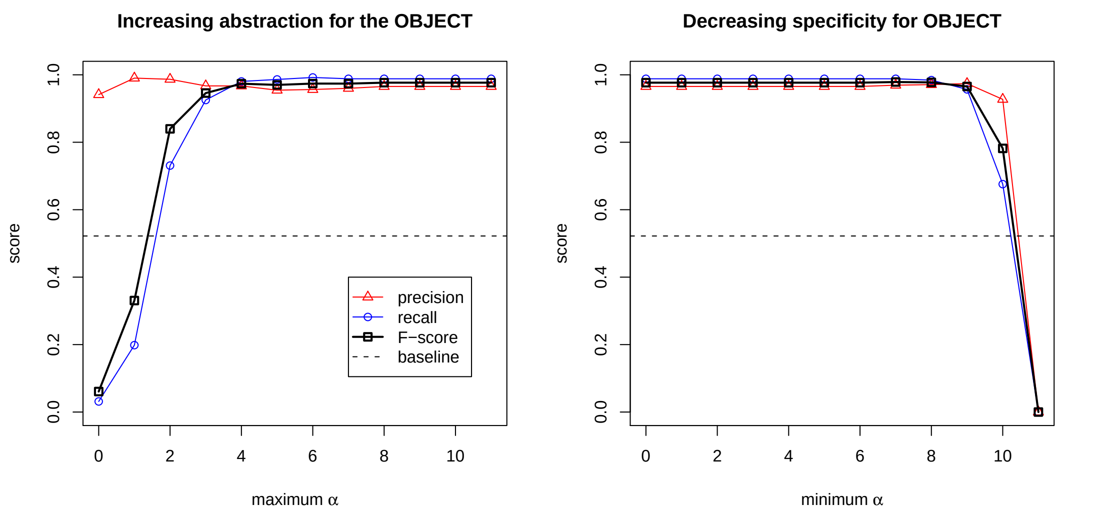


Figure 5.1: Scores for the `object` relative clauses on the dimensions of increasing abstraction and decreasing specificity.

		Maximum α											
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Minimum α	**0**	0.06	0.31	0.60	0.84	0.94	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	1		0.31	0.60	0.84	0.94	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	2			0.60	0.84	0.94	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	3				0.84	0.94	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	4					0.94	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	5						0.97	0.97	0.97	0.97	0.97	0.97	0.97
	6							0.97	0.97	0.97	0.97	0.97	0.97
	7								0.97	0.98	0.98	0.98	0.98
	8									0.98	0.98	0.98	0.98
	9										0.97	0.97	0.97
	10											0.82	0.82
	11												0.00

Table 5.11: F-scores for the `subject` relative clauses at different ranges of abstraction.

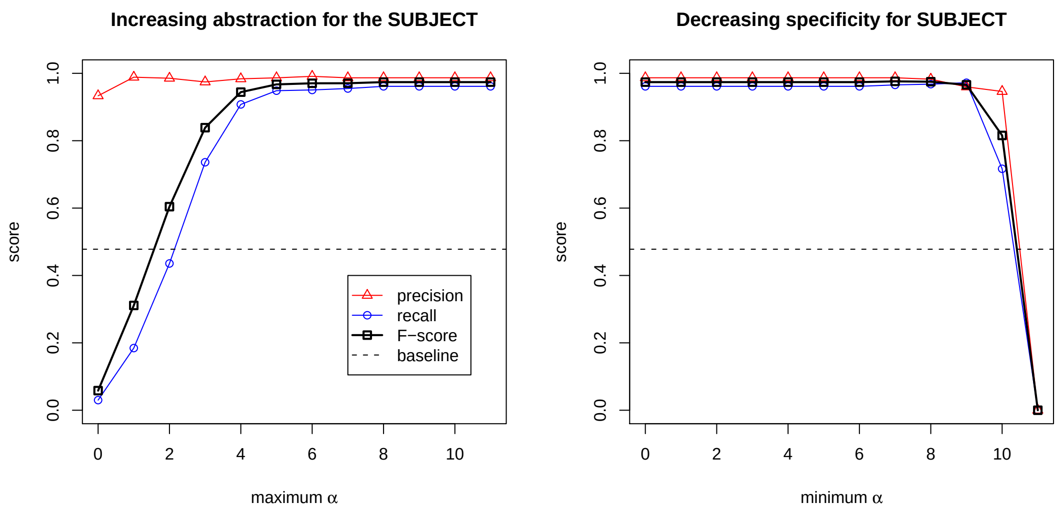


Figure 5.2: Scores for the `subject` relative clauses on the dimensions of increasing abstraction and decreasing specificity.

5.4 Qualitative exploration of the models

5.4.1 Global interpretation

In predicting the interpretation of transitive relative clauses, AM performs outstandingly. In many of the ranges of abstraction, F-scores of $F = 0.978$ are achieved for both interpretations. Given these features, AM can thus be said to be a good model of disambiguating transitive relative clauses.

When comparing the different ranges of abstraction, there seems to be no clear optimum. The model that allows for all supracontexts, $R_\alpha = [0 \ldots 11]$, reaches an F-score of $F = 0.974$ on the `subject` clauses and $F = 0.977$ on the `object` clauses. As we can see in the right panels of the two figures, decreasing the allowed specificity only causes a quick drop in the accuracy at $R_\alpha = [10 \ldots 11]$, meaning that only supracontexts with one or zero values specified are allowed. For the values in between, we can see that there the accuracy scores hardly decrease with an increasing minimum allowed level of abstraction α. Given the fact that all features point so strongly in the same direction, combining them seems not necessary to make good predictions. The model is still able to predict novel cases with highly abstract, rule-like schemas.

Increasing the allowed degree of abstraction of supracontexts also has little effect. The left panels show how both curves quickly reach their asymptote ($R_\alpha = [0 \ldots 4]$ for `object` and $R_\alpha = [0 \ldots 5]$ for `subject`. After this, the scores remain highly stable. Perhaps somewhat paradoxically, in the light of the fact that decreasing specificity does not harm the model, this means that beyond $R_\alpha = [0 \ldots 5]$ no more abstraction is needed either. With schemas specifying minimally six values, the model performs equally well as with all schemas, or as with only those containing maximally two features. Because all predictors are so strong individually, this off-line experiment cannot tell us much about the actual abstraction in the schemas of the transitive relative clauses. It hardly excludes any ranges of abstraction, except for the extreme ones, which more or less follows from the nature of Analogical Modeling.

5.4.2 An example

Let us begin with an analysis of how the outcome of a new item is determined. Take the sentence in (5.12), which can be coded as the exemplar in (5.13).

(5.12) *dan is er toch geen selectie vooraf geweest die enig niveau biedt ?*
than is there PRT no selection beforehand been that any level offers ?

'In that case there hasn't been any selection that provides some quality beforehand, hasn't there?' **(fn000746.158)**

(5.13) $\langle$`selectie quant.np abstract singular niveau quant.np abstract singular bieden singular NP`$\rangle$

supracontext	α	`object`	`subject`
diverse supracontexts with less than 10 exemplars	4	0	1
diverse supracontexts with less than 10 exemplars	5	64	46
⟨`- quant.np abst sing - - - sing - sing -` ⟩	6	100	0
⟨`- quant.np abst - - - - sing - sing arg` ⟩	6	196	0
diverse supracontexts with less than 10 exemplars	6	128	256
⟨`- - - - - quant.np abst sing - - arg` ⟩	7	0	121
⟨`- - - sing - quant.np - - - sing arg` ⟩	7	0	121
⟨`- quant.np - - - - abst sing - - arg` ⟩	7	0	324
⟨`- quant.np abst sing - - - sing - - -` ⟩	7	100	0
⟨`- quant.np abst sing - - - - - sing -` ⟩	7	100	0
⟨`- quant.np abst - - - - - - sing arg` ⟩	7	196	0
⟨`- quant.np abst - - - - sing - sing -` ⟩	7	784	0
diverse supracontexts with less than 10 exemplars	7	81	502
⟨`- - - - - quant.np - - - sing arg` ⟩	8	0	121
⟨`- - - - - quant.np abst - - - arg` ⟩	8	0	144
⟨`- - - - - quant.np abst sing - - -` ⟩	8	0	225
⟨`- - - sing - quant.np - - - sing -` ⟩	8	0	225
⟨`- - - - - quant.np - sing - - arg` ⟩	8	0	289
⟨`- quant.np - - - - abst - - - arg` ⟩	8	0	324
⟨`- quant.np - - - - abst sing - - -` ⟩	8	0	841
⟨`- quant.np abst sing - - - - - - -` ⟩	8	144	0
⟨`- quant.np abst - - - - - - sing -` ⟩	8	784	0
diverse supracontexts with less than 10 exemplars	8	0	332
⟨`- - - - - quant.np - - - sing -` ⟩	9	0	225
⟨`- - - - - quant.np abst - - - -` ⟩	9	0	361
⟨`- - - - - quant.np - sing - - -` ⟩	9	0	441
⟨`- quant.np - - - - abst - - - -` ⟩	9	0	841
diverse supracontexts with less than 10 exemplars	9	0	25
		2677	5765

Table 5.12: Homogeneous supracontexts for the given context in 5.13 supported by 10 or more exemplars. The abbreviations `quant.np` stands for quantified indefinite NP, `abst` for abstract and `sing` for singular.

Now, suppose this sentence is heard by a speaker after having encountered all other relative clauses in the present dataset. Based on the given context, 96 homogeneous supracontextual comparisons can be made. The set of all homogeneous supracontexts is then constituted by all supracontexts displayed in table 5.12. Many of these are based on only one exemplar, so we will display only the cases in which ten or more exemplars support a schema.

As we can see, the majority vote at $R_\alpha = [0 \ldots 11]$ correctly predicts the interpretation of this clause to be `subject`, that is: the selection provides the quality and not vice versa. Nevertheless, the prediction of the interpretation of this case goes wrong at several ranges. If we only allow for supracontexts with maximally four open slots ($R_\alpha = [0 \ldots 4]$), the model bases the prediction on only one exemplar that shares everything except for the word lemmas and the structure of the antecedent (*een zaak die wat tijd nodig heeft* ‘a case that needs some time’). This is a `subject` interpretation, so the nearest neighbor predicts the correct outcome.

However, when we allow for some more abstraction, the prediction goes wrong. As the table shows, among the supracontexts with five or six open positions, the most frequent outcome is the `object`-interpretation. Allowing for seven open positions ($R_\alpha = [0 \ldots 7]$), finally, causes the model to make the right prediction again, mainly because the extremely strong supracontextual schema $\langle$ `- quant.np abst - - - - sing - sing -` $\rangle$ is allowed. Many more abstract patterns that also predict the `subject` reading are equally strongly entrenched, so that a reduction of the specificity has no influence on this case.

In constructionist terms, the disambiguation between a `subject` and an `object` interpretation amounts to integrating the new object and its immanent schemas with one of the networks. In many cases, this disambiguation is easy and there is no reason for the cognitive agent to doubt. In this case there is, as nearly half of the votes is in favor of the `object` outcome. There are thus parts of the networks of both constructions pulling on this novel item. Table 5.12 lists the strongest supracontexts, but the link with the constructicon becomes clearer when we visualize these strongest schemas as nodes in a network. We can see how these nodes are not just disparate schemas, but coherent clusters, revolving around a set of properties they share by means of family resemblance. Figures 5.3 and 5.4, at the end of this chapter, give the strongest clusters for the two patterns.

For the `subject` readings, we can see that the leftmost cluster revolves around quantified NP, singular antecedents and abstract arguments. The second cluster has quantified NP, abstract arguments as relative-clause internal arguments. Note that the combination of these factors does occur as well, but in so few exemplars that they matter little in the prediction.

The rightmost cluster also contains some of these features, but moreover the aspect of agreement between the antecedent NP and the finite verb in the relative clause. The features most prominent in the prediction of this case as a `subject` are thus the empathy and structure of the antecedent and argument, as well as the agreement between the antecedent and the finite verb. Throughout all constructions, the feature `word order = arg` is found.

The `object`-network, displayed in figure 5.4, contains one big cluster of interrelated nodes. Central to these clusters is the abstract empathy level of the antecedent and the fact that it has a quantified NP-structure. Abstract antecedents often occur with an animate argument that functions as the subject.

Also in the `object`-constructions, agreement plays a role. In several constructions, the singular number of the finite verb as well as of the argument is present, which amounts to saying that these two agree.

Summarizing, by using the network structure that is immanent in the supracontextual schemas, we can gain insight in the features that favor predictions for either category.

Another important feature of the Analogical Modeling approach to classification is that the prediction does not rely on a single feature. This is a strength of the model in a case like 5.14. As the finite verb agrees numerically with the pronominal argument *dat* and not with the plural antecedent *mensen*, a model relying solely on the formal feature of agreement would predict that the relative pronoun would be the direct object of the main verb *doet* 'does'.

(5.14) *d'r zijn dus meer mensen die dat doet .*
there are PRT more humans that that does .

lit: 'So there are more people who does so.'

In real life, however, communication does not break down and errors such as these are probably more common than we think they are. Despite the fact that the exemplar displays agreement between the argument and the verb and not the antecedent and the verb, the model predicts that the relative pronoun is the subject of the clause, because everything else points to this interpretation: the combination of the antecedent *mens* and the argument *dat*, the animacy pattern by which the antecedent is a human being and the argument is an abstract entity (because it refers to a sentential antecedent in a previous clause) and combinations of these with the verb *doen*. Relatively few exemplars form good analogues that predict an object reading for this case. The object reading was favored by $21,118$ votes againts $228,784$ votes for the subject reading, in the range $R_\alpha = [0 \dots 11]$. Furthermore, at none of the other ranges, the prediction went wrong. The model thus is dynamic enough, as humans are, to deal with cases in which a formal pattern that predicts an outcome with near certainty, points in the wrong direction, relying on functional information in the clause.

5.4.3 Modeling the pyscholinguistic findings

So far, we have been making the idealization that only after all variables have been processed, the language user forms the interpretation. The studies of reading times in relative clauses, and of processing language in general, show that such an account of processing language is plainly wrong, and that an interactant's hypotheses about the interpretation are set up linearly and incrementally, as the sentence is processed. In this section, I show how AM can be used to mimick such findings, on the basis of Mak, Vonk & Schriefers's (2002) results on the processing of relative clauses.

Mak, Vonk & Schriefers (2002) found in reading and eye-tracking experiments that it is not so much the property of being an object relative clause that makes some relative clauses harder to process, but the fact that there is true semantic ambiguity in the clause, that has arisen through the fact that both participants are animate. This effect was shown by letting experimental subjects read sentences from the following four conditions, as exemplified in sentences (5.15-5.18), adopted from Mak, Vonk & Schriefers (2002, 56)

1. animate antecedent, animate argument, subject clause;

2. animate antecedent, animate argument, object clause;

3. animate antecedent, inanimate argument, subject clause;

4. inanimate antecedent, animate argument, object clause;

(5.15) ... *de inbrekers, die de bewoner beroofd hebben* ...
... the burglars, that the occupant robbed have ...

'The burglars, who robbed the occupant'

(5.16) ... *de bewoner, die de inbrekers beroofd hebben* ...
... the occupant, that the burglars robbed have ...

'The occupant, whom the burglars robbed'

(5.17) ... *de inbrekers, die de laptop gejat hebben* ...
... the burglars, that the laptop stolen have ...

'The burglars, who stole the laptop.'

(5.18) ... *de laptop, die de inbrekers gejat hebben* ...
... the laptop, that the burglars stolen have ...

'The laptop, that the burglars stole.'

If we expect that the object clauses are harder to process, the conditions 2 and 4 would both be harder to process than 1 and 3, as the former two are object clauses and the latter two are subject clauses. This effect is expected to show up at the processing of the auxiliary *hebben*, where the formally disambiguating number information is given, and the default hypothesis that the relative clause has a subject reding has to be overthrown.

Mak and colleagues, however, formed the alternative hypothesis that the processing difficulty of condition 2 arises not so much because of it being an object clause, but because there is no difference in animacy, so that it is harder for the language user to integrate the information. This difficulty is absent if either one of the two participants is ranked higher on the empathy hierarchy than the other, or if the antecedent is an animate entity and the relative clause has a subject reading.

Mak et al. claim that if the antecedent is animate, the expectation is that the relative clause will be a subject clause, but that this is not the case with inanimate antecedents. For this latter group, the expectation is that the relative clauses will have an object reading, as inanimate entities are more likely to be the direct object of a transitive pattern than the subjects. The expected effect is thus that the processing difficulty will arise in condition 2, but not in the other three conditions.

This hypothesis was confirmed in a reading-time experiment. For condition 2, but not for the other conditions, a significantly higher reading time was found at the auxiliary, presumably because the auxiliary is, for condition 2, the place where the default hypothesis (that given an animate antecedent the clause has a subject interpretation)

section contents	A antecedent	B argument	C participle	D auxiliary
variables	`ant` `structAnt` `empAnt` `nAnt`	`arg` `structArg` `empArg` `nArg`	`word order` `verb`	`nVerb`
example	*de inbrekers, die*	*de bewoner*	*beroofd*	*hebben*

Table 5.13: The four sections in the sentence and the features processed at that point.

has to be overruled by the language user because the agreement in grammatical number is known at that point.

This experiment can be simulated in AM by giving the model only the set of variables processed up to that point, so that it knows more or less as much as the incrementally processing human agent does. The other variables were marked as dummies, so that the model did not recognize them as any familiar value. What the model does, then, is generating the best sets of analogues, given the incomplete set of variables up to that point, but drawing on the memorized set of fully-processed schemas in memory. As such, we mimick the use of constructional schemas in incremental processing.

The sentences were split up in several sections. For each section, we ran the exemplar through `raam` with a manual mode, on the basis of the data set used in the rest of this chapter. The sections of the sentence, and the variables that go with it, and the example of condition 1 are shown in table 5.13.

The four sentences above were used as test items for the four conditions, and were made into the given contexts given in (5.19-5.22) The reader familiar with Mak and colleagues' work will have noticed that two lexical items have been changed. Whereas Mak, Vonk & Schriefers (2002) used the verb *stelen* 'steal', I used *jatten*, a verb with the same meaning but somewhat more colloquial. This was done, because some exemplars in the training set contained this verb and I wanted to exclude pure lexical effects. The same was the case with *computer* 'computer', which was replaced by *laptop* 'laptop'. The other lexical items used in the experiment were not found in my training set.

(5.19) ⟨ `inbreker def.np human plural bewoner def.np human singular beroven plural arg` ⟩

(5.20) ⟨ `bewoner def.np human singular inbreker def.np human plural beroven plural arg` ⟩

(5.21) ⟨ `inbreker def.np human plural laptop def.np object singular jatten plural arg` ⟩

	section			
condition	A	B	C	D
1	95%	87%	73%	99%
2	84%	100%	80%	25%
3	95%	100%	100%	100%
4	2%	6%	4%	1%

Table 5.14: The percentages of votes for the `subject`-intepretation for the four test sentences, per section.

(5.22) ⟨ `laptop def.np object singular inbreker def.np human plural jatten plural arg`⟩

What AM gives us per section processed, is a number of votes for each of the outcomes. We can transpose this into a percentage score for the subject reading, so that 8 votes for the `subject` reading against 2 for the `object`-interpretation will mean an 80% score for the `subject` interpretation. What we expect to happen is that if the sentence of the second condition is indeed harder to process, there will be some change in the interpretation-hypothesis as made by the model in condition 2 at section D, when the numerical agreement information is processed, and not in the other cases.

The four test items were processed as described and the scores per section per test item can be found in table 5.14. For the experiment the setting $R_\alpha = [0 \ldots 11]$ was used, as it turned out that the diffferent settings of abstraction had little effect.

These results are informally and rather unsystematically gatherered, but nevertheless are fully consonant with the findings in the first experiment of Mak, Vonk & Schriefers (2002). When processing the numeric information of the condition-2 clause, the model will have to change its hypothesis from `subject`-interpretation to `object`-interpretation. Moreover, the change is rather drastic: whereas after processing section C the outcome `subject` was favored for 80% in condition 2, this number drops to 25% in section D, a drop of 55%. If, at this point the reading times are significantly slower than in the other conditions, this comes as no suprise for AM. The model predicts that there is a drastic change in interpretation at section D of condition 2, but nowhere in any of the other conditions.

What this analysis furthermore shows us, is that in cases where there is difference in the animacy of the participants, such as conditions 3 and 4, there is hardly any ambiguity in the interpretation, and the certainty of the chosen interpretation is always well above 90% throughout the processing. For condition 2, as we have seen, the interpretation changes at section D, but for condition 1, some ambiguity arises as well, with at section C only 73% of the votes being in favor of the `subject`-interpretation. This effect does not surface in Mak et al.'s data, but it would be interesting to see whether there are other effects noticable of the fact that the degree of ambiguity is higher for condition 1 at section C, than for condition 3 at section C.

5.5 Conclusion

How to interpret the distribution of the models' performance at different ranges of abstraction? Based on the scores, we can formulate three subquestions:

- How to interpret the fact that at ranges only including very concrete supracontexts, the model performs poorly?
- How to interpret the fact that at ranges only including very abstract supracontexts, the model performs poorly?
- How to interpret the fact that most models in between of these extremes perform equally well?

The ranges containing only highly concrete supracontexts (minimum $\alpha \geq 2$) perform poorly because many cases cannot be predicted. This, in turn, is because these cases deviate too much from other exemplars to be captured by a schema with maximally two slots. The problem here is that of generalization: the specific constructional knowledge derived from experienced exemplars is insufficiently 'open-ended' to be used in the categorization of new cases.

This effect needs to be considered carefully, as we have seen it in the previous chapter as well. In cognitive modeling, we must always consider the possibility that results are artefacts of the model under certain conditions. I believe this generalization problem is such a case. AM is a categorical model: an exemplar falls within a supracontext or not and there is no such thing as 'distance' in terms of features, as there is in other exemplar models such as MBL (Daelemans & van den Bosch 2005). This causes the model to be simply unable to predict new cases if the supracontexts of these cases are empty. However, the only times no non-empty homogeneous supracontexts are available, is when we set the range of abstraction to one of its extremes.

But introducing the fuzziness of actual categorization into a categorical model from which fuzziness is only expected to *emerge*, would cause an undesired hybrid, as the source of any fuzziness in the categorization is no longer clear.

Because we have a clear indication of one source of the poor performance of the model at high minimum levels of α, we can say that these results are at least partly due to the nature of the model. Obviously, a comparison with the other experiments can shed more light on the performance with only concrete information. Still, very little abstraction is needed to make the model perform at its optimum. From $R_\alpha = [0 \ldots 4]$ onwards, both categories can be predicted nearly perfectly. This means that little abstraction is needed to predict this construction. Schemas with minimally seven values specified, or four open, suffice.

As to the second question, we also must consider the possibility of some scores being an artefact of the model as well. Supracontexts with few specifications often have high numbers of exemplars underlying them. There are, after all, many exemplars containing just a human antecedent and an agreeing argument $\langle$- - - hum - - agr$\rangle$. Because of this and the fact that abstract supracontexts also contain many subcontexts, the probability of one of these subcontexts not behaving like the supracontext is relatively high, and relatively few abstract supracontexts are likely to be homogeneous.

If we let the model predict the outcome of a new case, based on only abstract supracontexts, the likelihood of finding any predictor thus becomes smaller for reasons mentioned above. Cognitively, the principle whereby only homogeneous supracontexts may predict the outcome may be too strict in this case. Even in the case where there is only one out of, say, sixteen subcontexts having less internal disagreements than the supracontext at hand, the supracontext is considered worthless in the model. I thus interpret the fact that few supracontexts with an extremely low range R_α can be used in predicting the outcomes to be an artefact of the model. Nevertheless, it remains to be seen in other data sets how abstract models perform.

Now, if the extreme models do not tell us much about categorization in grammar, does the group of models with ranges of schematicity in between provide us with insightful information about the possible cognitive reality behind the categorization of transitive relative clauses?

I believe the insight this specific experiment gives us in the abstraction question is limited. The prediction relies on many factors in case of the interpretation of transitive relative clauses, but individually these factors are already good predictors of the outcome. By itself, the animacy of the two arguments, the word order in the relative clause, case-marked pronouns and the agreement pattern are each outstanding predictors. The data set is thus relatively regular, and for few cases did we truly find competition between the outcomes. Because of this regularity, it does not matter whether a language user stores it abstractly or in very concrete patterns as they all predict the same. In a less regular data set, we expect multiple competing abstract patterns that are sometimes overruled by highly specific supracontexts specifying exceptions to the rules. Such cases were hardly found in this dataset.

For this experiment, it all depends on what one likes to see: if the avoidance of redundancy is a virtue, the abstract model $R_\alpha = [8 \ldots 11]$ is a perfect model, if one favors the claim that no abstractions beyond the necessary ones are to be made, the model given $R_\alpha = [0 ldots 4]$ is the best candidate. Given the dictum of Langacker (2009), that all abstraction is immanent in the processing of the exemplar, this experiment does not decide between any of the models that perform equally well.

Does this result then not invalidate the model in general? I believe not. In fact, I believe it aids us in the interpretation of the results of the other experiments. Other curves, such as that of the PP-dative, become more interesting in the light of these results. If results deviate strongly from the results obtained in this experiment, they have more significance, and it is less likely that they are an unwanted effect of the mechanisms of the model. Furthermore, we have shown that human beings can perfectly well process and categorize the interpretations of relative clauses with an analogical, exemplar-based model. Not only does the categorization proceed as expected, also the processing of these clauses, in terms of reading times, can be predicted using an analogical and constructionist model.

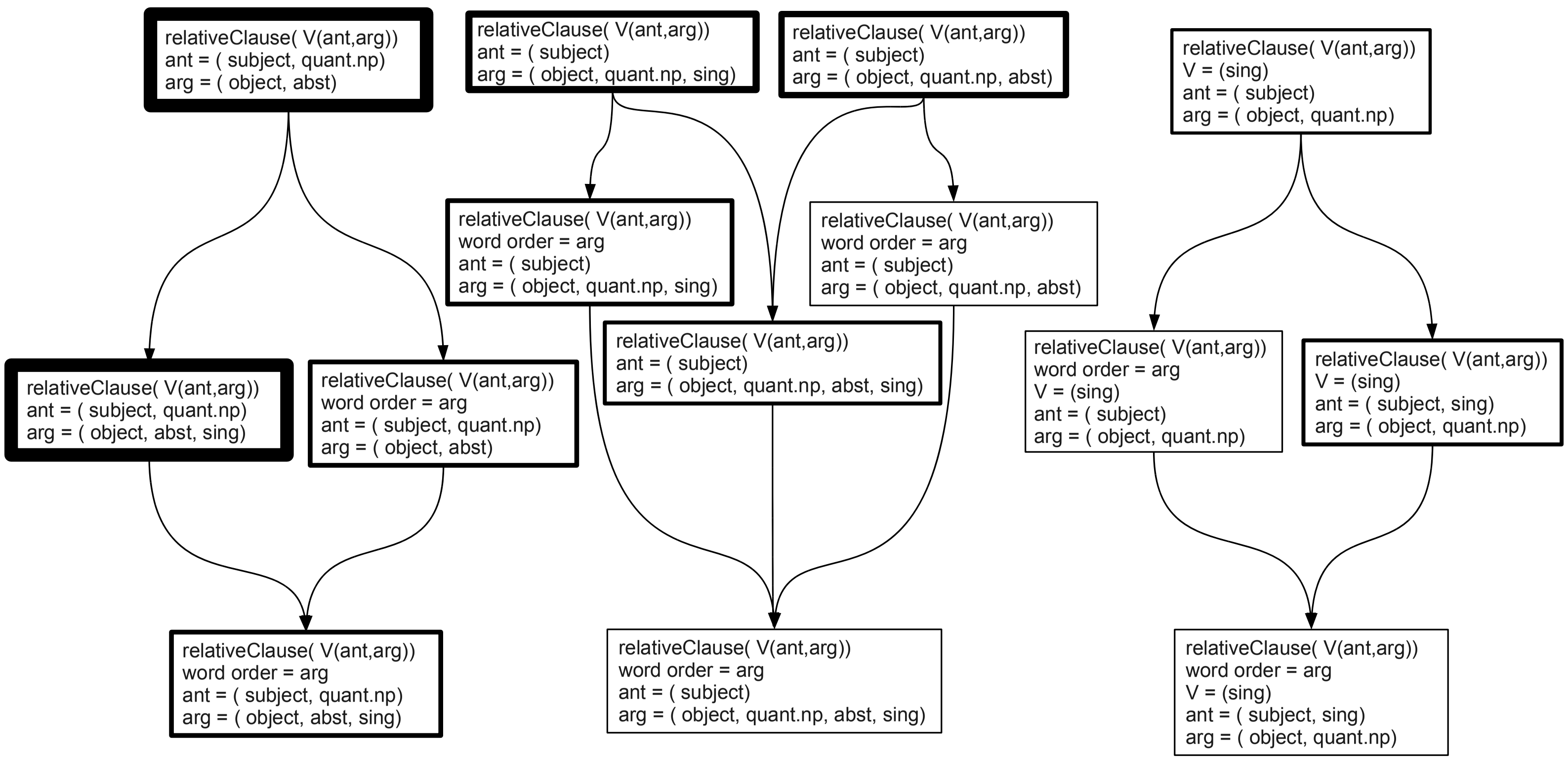


Figure 5.3: The strongest supracontexts predicting the `subject`-reading. The supracontextual schemas are presented in a pseudo-constructionist way. The thickness of the boxes resembles the entrenchment, with one pixel standing for fifty votes.

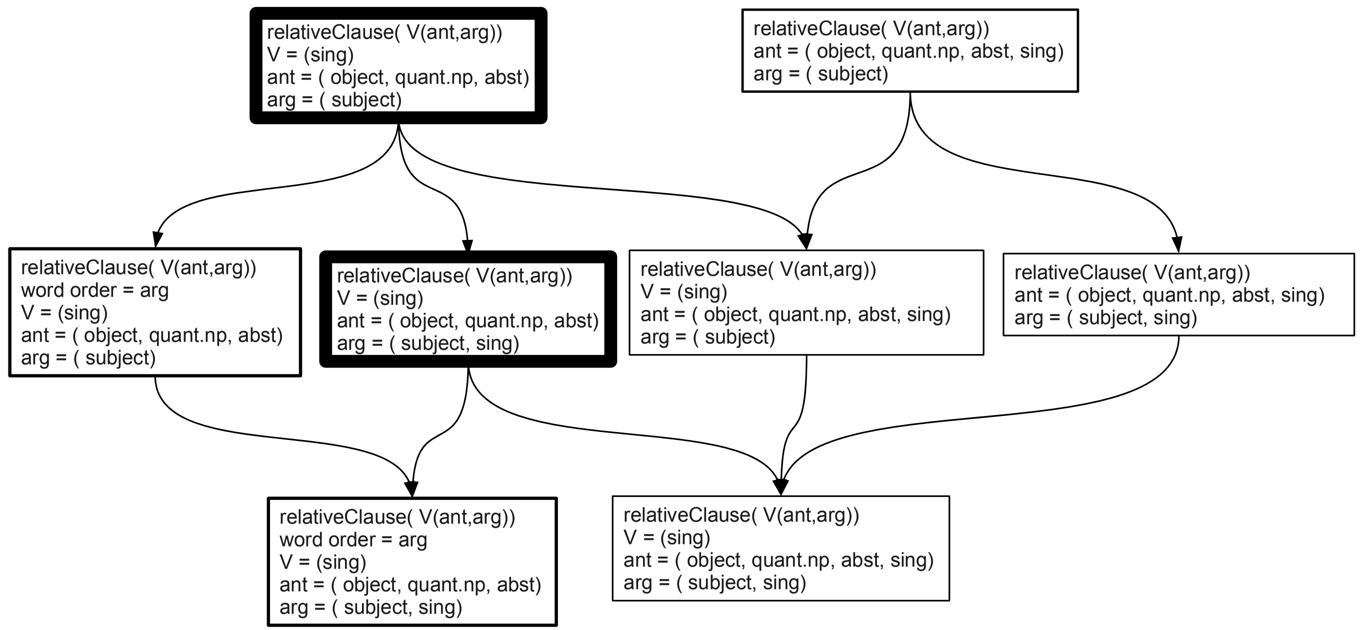


Figure 5.4: The strongest supracontexts predicting the `object`-reading. The supracontextual schemas are presented in a pseudo-constructionist way. The thickness of the boxes resembles the entrenchment, with one pixel standing for fifty votes.

Chapter 6

A multivariate grammar: Dutch progressive constructions

6.1 Introduction

We have seen how we can model the categorization of grammatical patterns in the chapter on the dative alternation and how the categorization of the semantic roles in transitive relative clauses proceeds in a similar manner. For several reasons, it is desirable to supplement the findings from these two studies with a third one. First of all, the aim of this thesis is to explore the method of investigating abstraction. With only four categories studied so far, the scope of application of the model is rather limited and few general claims can be made. To supplement this, we use a categorization task involving five categories, namely the constructional choice of progressive marking.

A second reason for executing a third experiment is that all variables in the previous two chapters were derived from approximately the same two domains, namely form and the semantic construal of the situation. In this chapter, I model a phenomenon in which we use variables from other domains, to show how such a multivariate approach can be executed naturally with an exemplar-based model.

6.2 The constructions and their determinants

6.2.1 Expressing progressive aspect in Dutch

In Dutch, progressive aspect is optionally marked with one of five constructions. The marking is typically optional because an imperfect past or present may convey the continuation of a situation as well, with the exception of the lexical-aspectual category of achievements presented as ongoing activities (Boogaart 1999, 191-192). In the present case study, the category of imperfects is not included, as it does not mark progressive aspect explicitly. The five patterns used for marking the progressive that are included fall into three classes.

First, there are three cardinal position verbs that can be used as auxiliaries to construe a situation as ongoing (sentences (6.1)-(6.3)). The structure of this construction consists of a cardinal position verb as an auxiliary, and the main verb as an infinitive, preceded by the preposition *te* 'to, at'.[1] As there are major differences between the use of the three cardinal posture verbs, I will use each of them as a separate category in the classification task. The labels used are `zit` for the pattern with *zitten* 'sit', `lig` for the pattern with *liggen* 'lie' and `sta` for the pattern with *staan* 'stand'. Interestingly, this group of cardinal posture verbs can be marginally expanded with other verbs, such as *hangen* 'hang'. A structurally similar pattern that was not taken into account in this study was the idiom *Dat staat je te wachten* (lit: 'It stands you to wait'; 'you will experience that in the future') and variations, as this type has no counterpart without *staan* and the other progressives are not even remotely acceptable with this meaning (**Dat zit je te wachten/*Dat is je aan het wachten*; lit: 'It sits you to wait/It is jou on the wait')).

(6.1) *en uh en Jim die lag ook te slapen .*
and uhm and Jim that lay also to sleep.INF .

'And Jim was also sleeping.' **(fn007980.81)**

(6.2) *en d'r staat er eentje te paffen .*
and there stands there one.DIM to smoke.INF .

'And there is one smoking over there.' **(fn008473.160)**

(6.3) *en Niels zit echt superraar te eten .*
and Niels sits really super.strange to eat.INF .

'And Niels is eating in a really strange way.' **(fn000554.378)**

Then, there is the possibility of marking the progressive with the verb of motion *lopen* 'walk' and the main verb in the infinitive, preceded by the preposition *te* 'to', again with *lopen* in the infinitive and *te* dropped if there is another auxiliary. An example of the pattern can be found in example (6.4). Again, the category of motion verbs to express progressive grammatical aspect is marginally extendable with 'creative' extensions, such as *hij rijdt te bellen* lit: 'he drives to call', 'he drives and calls', but the pattern with *lopen*, henceforth the `loop`-pattern is the only conventional one.

(6.4) *maar ja liep die weer te zeiken .*
but yes walked that again to piss.INF .

'But yeah, he was nagging again' **(fn000702.183)**

[1] Except for cases in which yet another auxiliary is used. In that case, the posture verb is in the infinitive form and the preposition *te* is dropped: *Hij heeft staan wachten*, lit: 'He has stand.INF wait.INF', 'He has been waiting'.

Then, progressive aspect can be marked with a construction involving a (copular) verb, the preposition *aan* 'on', and a nominalized infinitive with the article *het* or its reduced form *'t* 'the'. This category will henceforth be called `ah`, short for *aan het*, the two fixed morphemes of this pattern. Sentence (6.5) gives an example. Note that different verbs can function as the auxiliary verb in this construction, but that many of them do not involve mere progression. Booij (2008, 85-85) lists the following categories:

1. Copular verbs that express appearance, such as *blijken*, *lijken*, *schijnen* 'appear to be'. An example is *Hij lijkt aan het eten* 'He seems to be eating';
2. Verbs that express perception: the so-called accusativus-cum-infinitivo verbs, such as *horen* 'hear', *zien* 'see'. An example is *Ik hoor hem op zolder aan het rommelen* 'I heard him pottering around in the attic';
3. Verbs that take a secondary predicate (*hebben* 'have' and *hold* 'houden'), an example being *Hij heeft de motor aan het lopen* 'He has the engine running';
4. Inchoative and continuative verbs, such as *gaan* 'go', *raken* 'get', *slaan* 'hit' and *blijven* 'stay, remain'. An example is *Ze slaan aan het knokken* 'they start fighting';
5. Causative verbs, such as *brengen* 'bring, *maken* 'make, *krijgen* 'get and *zetten* 'to put. An example is *De politie kreeg hem aan het praten* 'The police got him talking'.

In this experiment, I only included those categories that have a counterpart in the other progressive constructions, viz. the inchoative and continuative verbs and the epistemic copular verbs of appearance (see sentences (6.6) and (6.7)). The other three categories were excluded because they have no direct counterpart in the other progressive constructions (see sentences (6.8)-(6.10)), which reflect my intuition and were not attested), except possibly for the accusativus-cum-infinitivo verbs, which are marginally grammatical with the `zit`-pattern in my intuition but which were not found in the corpus with any categories but the `ah`-pattern. The three excluded categories furthermore have a meaning that modifies the participant roles of the predicate: the accusativus-cum-infinitivo verbs contain an additional experiencer and the secondary predicate verbs and the causative verbs contain an additional causer or possessor.

(6.5) *we zijn echt de hele tijd aan 't praten .*
we are really the whole time on the talk.INF .

'We are really talking all the time.' **(fn000333.305)**

(6.6) *We gaan (zitten te / aan het) denken over een oplossing .*
We go (sit to / on the) think.INF about a solution .

'We'll start thinking about a solution.'

(6.7) *Ze bleken (aan het eten te zijn / te zitten te eten) .*
They appeared (on the eat.INF to be / to sit at eat.INF) .

'They appeared to be eating'

(6.8) *We hebben de motor (aan het / *zitten te) lopen .*
We have the engine (on the / *sit.INF to) running .

'We have the engine running.'

(6.9) *Ik zag hem (aan het / ?zitten te) wachten .*
I saw him (on the / ?sit.INF to) wait.INF .

'I saw him waiting.'

(6.10) *Ik kreeg hem (aan het / *zitten te) praten .*
I got him (on the / *sit.INF to) talk.INF .

'I got him talking.'

It may seem that the four categories involving a cardinal position or motion cannot be compared with the seemingly more grammatical or abstract `ah`-pattern, because they still have a lexical meaning and are as such not always acceptable where a case of the `ah`-construction is. As Lemmens (2005) and Booij (2008) argue, the lexical meaning is at least backgrounded and is in many cases absent or irrelevant (example (6.11)) or even incompatible (example (6.12)).

(6.11) *ja ik zat al de hele tijd te kijken .*
yes I sat already the whole time to look.INF .

'Yeah I was already looking all the time.' **(fn000293.416)**

(6.12) *heeft ie toch altijd zo'n pennetje in z'n hand waar ie mee loopt te*
has he PRT always such.a pencil.DIM in his hand where he with walks to
draaien . . .
turn.INF . . .

'And then he always holds a little pencil in his hand with which he's twirling . . . ' **(fn000981.176)**

6.2.2 Determinants and corpus data

In a contrastive study with the English progressive, Boogaart (1999, ch. 5) discusses several constraints on the progressive. First, there are restrictions on the Aktionsart of the verbs. According to Boogaart, the class of states (cf. Vendler 1957) is incompatible with the Dutch progressive, except in marginal cases in which the state is framed as someone being "busy being [something] or trying to look like [something]" (Boogaart 1999, 176). In combination with a progressive marker, the states are thus conceptualized as activities. Booij (2008, 84) goes even further and claims that "[t]he Dutch progressive construction is restricted as to the kind of verbs it allows: the verb should be an activity or an accomplishment verb (that is a durational verb); stative and achievement verbs are excluded." As we will see, this claim is too strong. Although most progressives have a main verb from the Vendlerian class of activities, followed by accomplishments, the two other classes are certainly not excluded.

In the light of the constraint that passives cannot be made progressive, while the undergoers of some gradual change (such as *getting old*, *decay* and *develop*) can be marked with a progressive, Boogaart (p. 179) formulates the restriction that the subjects of the Dutch progressive constructions are either agentive or the undergoers of some gradual change. Moreover, the four progressive constructions derived from verbs expressing human positions and activities (`zit`, `sta`, `lig`, and `loop`) require their subject to be an entity that can sit, stand, lie or walk. This means that these subjects are typically agentive, although not necessarily so, because the posture verbs are obligatorily used for marking the position of inanimate objects as well. As we will see, the difference in agentivity of the subjects is indeed one of the factors driving the alternation.

Vismans (1982), as cited in Boogaart (1999, 182 ff.), furthermore discusses as a constraint on the use of the cardinal posture verbs that the activity be uninterrupted. Boogaart disagrees, but acknowledges that "using a [cardinal posture verb] is strange for telic situations the endpoint of which could not, in principle, be reached within one occasion of sitting, lying etc.". Whether or not the situation is interrupted or not, is impossible to code in a corpus, as it takes too much interpretation on the part of the coder, however, markers of repetition and the proportions of verbs being telic show that this claim is essentially true.

A major, recent study that forms the starting point for the selection of determinants in this experiment is that of Lemmens (2005), in which the progressive constructions on the basis of the three cardinal posture verbs are discussed. According to Lemmens, the three cardinal posture verbs often differ from the `ah`-pattern by marking the cardinal posture of the subject. In the cases where the posture of the subject is "no longer at issue" and sometimes even incompatible, the cardinal posture verbs seem interchangable with the `ah`-pattern. As the *Algemene Nederlandse Spraakkunst* (Haeseryn et al. 1997, 973), the Dutch reference grammar, states, the meaning of these auxiliaries has often weakened to 'be occupied with'. Sentences such as these often indicate irritation and occur mainly in the spoken language. Lemmens agrees with this connotational and lectal characterization, but claims that the variation should be determined by more than just stylistic and connotational variation, and that the source of the subtle differences can be derived from the original lexical meaning of the posture verbs. As this study is a synchronic categorization task, the precise relations of the patterns to their source lexical verbs falls beyond the scope of this study, and these motivations will only marginally be discussed in the following. For the sake of this experiment, I treat the situation as if the five patterns were actually on par with each other, whereas this may actually be an idealization.

Firstly, the different posture verbs each have a different distribution of agents executing them. In Lemmens' newspaper data, the `lig`-pattern has concrete inanimate entities as subjects in 40.7% of the cases. The motivation for this lies in the fact that human beings and other animates have only a limited range of activities that can be executed while lying, whereas this range of activities is less limited for animates while sitting and standing.

Then, there seem to be strong collocations between the different constructions and verbs. Here too, the lexical meaning of the auxiliaries plays a major role. In Lemmens' data, *slapen* 'sleep' occurs most often with the `lig` pattern, whereas *lezen* 'read' is most often found with the `zit` pattern. Sleeping is an activity we typically do while lying, and reading is usually done sitting. From these concrete activities associated with certain positions, Lemmens derives certain classes of activities that are associated experientally with the postures expressed by the lexical variants of the auxiliaries. Although I think Lemmens confounds several parameters (experiental domain, agentivity, and Aktionsart) in his classification and I will use a classification more oriented on the experiental domain, the patterns that emerge from his classification are clear: `sta` is associated with verbs of motion and communication, `zit` with verbs expressing something that happens while remaining in one place and verbs of cognition, whereas `lig` is associated most strongly with verbs expressing rest and motion that co-occurs with lying. In a short discussion of the `loop`-progressive, Lemmens (p. 212) finds that most instantiations of this pattern have a main verb that expresses an activity involving physical motion, such as *voetballen* 'play football'. When no actual motion is involved, the pattern typically expresses repetition, often combined with a verb that displays a negative evaluation on the speaker's part.

Finally, Lemmens discusses how durative and locative markers may shed light on the distinction between the cardinal posture verb patterns and the `ah`-construction. He finds slight differences between the three cardinal posture verb patterns on the one side and the `ah`-pattern on the other, but finds these to be too small to be interesting for a contrastive characterization. For the oblique locational markers, the results are stronger: whereas in 44% of the cases of the cardinal posture verb progressives, a locative marker can be found, only in 12% of the `ah`-progressive a locative marker is present. This difference can be attributed, according to Lemmens, to the fact that the cardinal posture verbs appear to require a locative complement when not used as an auxiliary ($^{?}$*Hij zit* 'He sits' vs. *Hij zit op de stoel* 'He sits on the chair'). An interesting finding of Lemmens is that the `loop`-progressive often marks repetition (as the lexical meaning does so as well), and that therefore more durational and aspectual modifiers are found in this pattern.

These diverse remarks make a good starting point for an investigation into the alternation between the five patterns. Such a study has not been done yet systematically for spoken Dutch, and this experiment fulfils the double role of being a case study for my methodological goal, as well as a description of a yet unthreaded part of spoken Dutch.

For this study, I extracted all cases of the five discussed patterns from components *a*, *f*, *g* and *h* of the Netherlandic Dutch section of the *Corpus Gesproken Nederlands*, which contained in total $2,818,823$ words. The extraction was done using regular expression queries with `grep` on the `.sea` source files of the corpus. This yielded a total of 2079 sentences, distributed over the five progressives as shown in table 6.1

Properties of the subject As Lemmens (2005) noted, there are differences between the kinds of subjects found with the different patterns. It seems, however, that he confuses the agentivity of the subject and its empathy level or animacy. For both, however, we can test to what degree they are distributed differently over the various progressive constructions. For the agentivity we can say that if a subject is non-volitionally participating in the situation denoted by the main verb, it will more likely be expressed with a cardinal posture verb progressive, as the lexical variants of these verbs denote an entity in a non-active position. Similarly, as the lexical variant of `loop` denotes the activity of walking, inactive participants are unlikely to be found with this pattern. As we can see in table 6.2, this pattern can be found in the data. I coded three types of subject-roles, namely `agent`, `subject` and `experiencer`. The cut-off point between the former two is the volitionality of the instigation and the continuation of the action. Thus, the subject of 'sleeping' is a thematic subject, the subject of 'yelling' is an agent and the subject of 'thinking' is an experiencer. Experiencers were coded separately, as experiencing seems to be a domain not easily captured in terms of actor or undergoer. Sentences (6.13)-(6.15) give examples of agents, experiencers and subjects respectively. This category will henceforth be called `agentivity`.

(6.13) *je bent nou hun aan 't afleiden .*
you are now them on the distract.INF .

'You are distracting them now' **(fn009066.36)**

(6.14) *maar zitten daar altijd nog wel lui te kijken dan Marnix ?*
but sit there always still PRT people to look.INF then Marnix ?

'But are there still people looking there, Marnix?' **(fn000756.257)**

(6.15) *'t gebied van overlap is op 't ogenblik aan 't groeien .*
the area of overlap is on the moment on the grow.INF .

'The area of overlap is currently growing' **(fn007255.9)**

Orthogonal to this distinction is that of the `empathy` level of the subject. Using the scale of Silverstein (1976), improved by adding the level of `organization` from Zaenen et al.'s (2005) coding schema, we can see that the different levels of empathy are not distributed evenly over the different progressive constructions. Most subjects are animate beings, either as `speaker`, `hearer`, third-person `human`, `organization` or as `animal`. For inanimates, the `sta` and `ah`-progressives are the most often used. Among the animate subjects, it seems that the two most frequent patterns,

	type of progressive					
	`ah`	`lig`	`lop`	`sta`	`zit`	**total**
frequency	753	48	90	220	968	2079

Table 6.1: Frequency of the progressive constructions in the used corpus.

	`agentivity`			
progressive	agent	experiencer	subject	**total**
`ah`	605	65	83	753
`lig`	10	4	34	48
`loop`	77	6	7	90
`sta`	137	41	42	220
`zit`	520	349	99	968
total	1349	465	265	2079

Table 6.2: Frequency of the progressive constructions per agentivity level of the subject argument.

`zit` and `ah` are also used frequently for speakers and hearers, whereas the other three have a preference for third persons. This category will be referred to as `animacy`.

Verbal properties Also certain properties of the main verb can be operationalized as variables in a categorization task. First of all, certain verbs seem to co-occur more frequently with certain progressive constructions than others. This is due to two reasons. First of all, the four progressives derived from lexical verbs are each associated with a specific set of experientaldomains. Sitting is associated with cognitive activities, ingestion and communication, whereas lying is associated with inactive situations such as resting (Lemmens 2005). Secondly, it may be due to mere idiomatization. Two additional factors used in the analysis thus are the `verb` and the experiental `domain`. The five most frequent verbs per category are shown in table 6.4.

For the domain of experience I based my coding practice on Lemmens' schema (Lemmens 2005, p. 201). However, Lemmens' coding schema seems to confuse agentivity, Aktionsart and domain of experience, and to tease out the effect of the experiental domain, I removed those categories that were not definable without reference to agentivity or Aktionsart. Examples of the categories are the following:

- `body`: body processes, such as 'sleeping', 'vomiting'

	empathy							
	speaker	hearer	human	org.	animal	object	abstract	**total**
ah	271	127	282	13	8	16	36	753
lig	10	6	24	0	5	3	0	48
loop	20	14	54	1	1	0	0	90
sta	62	19	110	2	9	17	1	220
zit	459	151	336	4	13	1	4	968
total	822	317	806	20	36	37	41	2079

Table 6.3: Frequency of the progressive constructions per empathy level of the subject argument. ‘org.’ stands for the empathy level of `organization`.

- `cognitive`: epistemic mental processes, such as ‘thinking’, ‘doubting’
- `communication`: communicative processes, such as ‘saying’, ‘complaining’
- `cultural`: frames being defined on a cultural level involving more than just physical actions, such as ‘economizing’, ‘writing a book’
- `emission`: processes of emission, such as ‘yelling’, ‘shining’
- `evaluation` evaluative mental processes, such as ‘enjoying’, ‘being fed up’:
- `generic`: verbs with vague semantics, sometimes denoting a manner in which an action is performed, such as ‘doing’, ‘being in a hurry’
- `ingestion`: processes of ingestion, such as ‘eating’, ‘smoking’
- `moving`: processes involving physical action while moving, such as ‘playing football’, ‘traveling’
- `perception`: processes of perception, such as ‘seeing’, ‘reading’
- `social`: frames being defined on a cultural level related social relations, such as ‘punishing’, ‘bullying’
- `stativeMotion`: processes involving physical, heterogeneous activity while remaining in one place, such as ‘typing’, ‘packing one’s bag’
- `stativePhysical`: processes involving physical inaction or homogeneous activity while remaining in one place, such as ‘waiting’, ‘blooming’

ah	lig	loop	sta	zit
doen 'do' (60) *praten* 'talk' (40) *kijken* 'look' (22) *vertellen* 'tell' (17) *zoeken* 'search' (15)	*slapen* 'sleep' (15) *janken* 'whine' (3) *lezen* 'read' (3) *creperen* 'die' (2) *wachten* 'wait' (2)	*zeiken* 'complain' (4) *zoeken* 'search' (3) 9 types with $n = 2$	*kijken* 'look' (31) *wachten* 'wait' (16) *praten* 'talk' (11) *goochelen* 'conjure' (7) *doen* 'do' (6)	*kijken* 'look' (119) *denken* 'think' (110) *wachten* 'wait' (55) *eten* 'eat' (33) *praten* 'talk' (31)
332 types	28 types	76 types	122 types	294 types

Table 6.4: Top five main verbs per progressive construction. The numbers between parentheses mark the absolute frequency. The bottom row gives the type frequencies per category.

	domain													
	body	cognitive	communication	cultural	emission	evaluation	generic	ingestion	moving	perception	social	stativeMotion	stativePhysical	**total**
ah	30	69	117	63	12	4	140	22	64	37	23	149	23	753
lig	22	1	0	0	6	2	1	0	0	4	0	8	4	48
loop	5	5	28	2	5	2	15	2	13	1	3	8	1	90
sta	9	4	37	0	19	5	8	6	7	37	8	50	39	229
zit	30	207	168	10	46	20	61	65	10	170	22	94	65	968
total	96	286	350	75	88	33	225	95	94	249	56	309	132	2088

Table 6.5: Frequency of the progressive constructions per experiental domain of the main verb.

Table 6.5 displays the domains I coded for and the distribution of the progressive constructions over them.

Then, we can wonder whether there is a division of labor between the progressive constructions in that some progressive constructions co-occur with certain Aktionsarten more often. We distinguish the four classes discussed by Vendler (1957), ameliorated with the class of `gradual completion` verbs (cf. Boogaart 1999, 179) and `semelfactive` verbs. The four Vendlerian classes are:

- `accomplisment`, being actions that have duration and that have an inherent ending point in a changed situation, such as 'eating an apple', where the apple is gone after the eating;
- `activities`, being actions that have duration, but that have no inherent ending point, such as 'running around', where the perceived situation is not changed after the agent decides to cut off the action;
- `achievements`, being actions that have no duration, and that do have an inherent ending point, such as 'arriving at the station', where the resulting state is that the agent is arrived at the station;

- `states`, being situations that have duration and no ending point, but that, unlike the other three classes, show no change in the situation. An example is 'having a car' or 'being a boy'.

Gradual completion verbs are a class on their own, as they have a inherent ending point that is gradually reached, such as 'getting older' or 'growing'. Semelfactive verbs are those that are punctual, but have no inherent result state that comes into effect through the action (e.g. 'knocking' or 'coughing'). A quick glance at the data (table 6.6) shows that the strong claims of Booij (2008), that verbs denoting achievements and states are incompatible with the Dutch progressive, are false. The more 'telic' Aktionsarten, viz. those that have either an inherent ending point or a gradually reached goal, display a preference for the `ah`-progressive, whereas the other two (`activity` and `state`) display more cases of the `zit`-progressive. The `loop`-pattern furthermore shows a relative preference for achievements, which can be expected from the repetitive meaning it developed. Finally, the high number of verbs expressing states with the progressive `sta` is remarkable, as Lemmens noted that this pattern has a very dynamic meaning due to the fact that standing is the starting point of many moving activities. The examples (6.16) through (6.20) illustrate cases of accomplishments, achievements, activities, states and gradual completion verbs respectively. Note that in the progressive, all the different Aktionsarten are conceptualized as activities. For instance, whereas *denken aan* 'thinking of' is a state in its base sense, unlike other cognitive, thinking-related verbs as *nadenken over* 'think about', it feels more like an activity when used in combination with a progressive marker. The variable of Aktionsart will henceforth be called `aktionsart`.

(6.16) *ik dacht Bas was een sjekkie aan 't draaien .*
I thought Bas was a rolled.cigarette on the roll.INF .

'I thought Bas was rolling a cigarette.' **(fn000559.192)**

(6.17) ... *wij staan hier een briefje te krijgen voor acht uur melden .*
... we stand here a note.DIM to get.INF for eight hours report.INF .

'We are getting a note stating that we should report ourselves at eight o'clock. **(fn000635.230)**

(6.18) *maar lopen die meiden ook zo te klieren dan ?*
but walk those girls also that.way to pester.INF then ?

'But are those girls making a nuisance of themselves as well then?' **(fn000723.227)**

(6.19) *hij zegt ik stond net j aan je te denken .*
he said I stood just j on jou to think.INF .

'He said: "I was just thinking about you"' **(fn007838.176)**

	aktionsart						
	accom-plishment	achieve-ment	activity	gradual completion	state	semel-factive	**total**
`ah`	115	28	511	69	29	1	753
`lig`	3	0	24	1	20	0	48
`loop`	4	6	72	3	3	2	90
`sta`	10	3	168	5	34	0	220
`zit`	42	31	656	7	228	4	968
total	174	68	1431	85	314	7	2079

Table 6.6: Frequency of the progressive constructions per Aktionsart of the main verb.

(6.20) *want die uh spinazie die staat te ontdooien hè .*
because that uhm spinach that stands to defrost.INF PRT .

'Because that uhm spinach, it is defrosting, right? **(fn000403.190)**

Next there is the claim of the *Algemene Nederlandse Spraakkunst* that there is often an undertone of irritation. Obviously, this is too subtle to operationalize directly, but cases that can be relatively simply marked as such are those in which the verb implies that the speaker evaluates the action negatively. Typically, these are verbs containing relatively little semantic content, such as *klieren*, *klooien* and *kutten*, all meaning 'to behave in an annoying manner', *kakelen*, *zeiken* and *ouwehoeren*, all meaning 'to talk in an annoying manner', with slight connotational differences between them. An attested example is sentence (6.18). As we can see, these types of verbs are most frequently found with `zit` and `loop`-progressives. This variable is called `evaluative`

Finally, we can investigate the presence of grammatical marking of aspect, that is through perfective auxiliaries. It seems that the `ah` pattern somehow disprefers occurring in the perfect, although a direct motivation for this fact has to be investigated further. The ingressive aspect, marked with *gaan* 'go', is also found disproportionally often with the `zit`-pattern. A non-imperfective aspect thus is a strong motivation not to use the `ah`-pattern. We will call the variable of grammatical aspect `aspect`.

Obliques Lemmens (2005) notices that the cardinal posture verb progressives often have oblique markers of location. In this study we coded the type of location expressed (if any) in the data, with the following levels:

- `object` referred to clear three-dimensional objects, such as 'on the table', 'in front of the door'

	evaluative		
	no	yes	**total**
ah	741	12	753
lig	47	1	48
loop	64	26	90
sta	217	3	220
zit	920	48	968
total	1989	90	2079

Table 6.7: Frequency of the progressive constructions per answer to the question whether the verb embodies the evaluative judgement of the speaker.

	aspect					
	none	ingressive	ingressive + perfective	imperfect	perfect	**total**
ah	0	9	1	736	20	753
lig	0	1	0	45	2	48
loop	1	7	0	61	21	90
sta	0	5	0	194	21	220
zit	0	52	2	789	125	968
total	1	74	2	1825	177	2079

Table 6.8: Frequency of the progressive constructions per type of grammatical aspect.

- space referred to physical spaces, often expressed with 'daar', but also cases like 'outside', 'in the attic'.
- abstract referred to non-physical, metaphorical spaces, such as 'in the economy'

In case of the absence of any locational markers, none was coded. The factor of type of location will be referred to from now on as location. Table 6.9 shows how the different categories are distributed over the location markers. The ah-pattern has the lowest amount of markers ($10.6\%, n = 90$), whereas a majority of cases ($60.4\%, n = 29$) of the lig-progressive mark the location explicitly.

Lemmens (2005) does not find significant differences between the presence of temporal and durative markers in the different progressive markers. Boogaart (1999, 182

	location				
	none	abstract	object	space	**total**
`ah`	673	12	3	65	753
`lig`	19	2	10	17	48
`loop`	76	1	2	11	90
`sta`	131	3	22	64	220
`zit`	817	12	36	103	968
total	1716	30	73	260	2079

Table 6.9: Frequency of the progressive constructions per type of meaning of oblique locational markers.

	aspectualMarking				
	none	precise duration	repetitive	vague duration	**total**
`ah`	653	16	21	63	753
`lig`	43	3	0	2	48
`loop`	71	4	4	11	90
`sta`	193	6	7	14	220
`zit`	778	43	27	120	968
total	1738	72	59	210	2079

Table 6.10: Frequency of the progressive constructions per type of meaning of oblique markers of aspect.

ff.) suggests, however, that the cardinal posture verb patterns seem to be more associated with actions that can be executed in one session of sitting, lying etc. In the light of this remark, we can expect that, if there are any durative markers expressing a precise amount of time, they are more frequently found with the cardinal posture verb patterns than with the `ah`-pattern. The variable of `aspectualMarking` was coded to capture this expectation. The levels of this feature are:

- `repetion`: items that mark the repetitiveness of an action, such as 'again', 'every day'
- `vague duration`: markers of duration without mentioning an amount of hours, minutes, days etc., examples being 'a long time', 'all the time'.
- `precise duration`: markers of duration mentioning an amount of time in hours, minutes, days etc., examples being 'from two to five o'clock', 'three minutes'.

	referencePoint						
	none	conditional	deictic	relative	precise	vague	**total**
ah	595	34	75	5	11	33	753
lig	36	3	2	0	5	2	48
loop	68	9	5	1	3	4	90
sta	167	21	7	5	9	11	220
zit	737	88	48	11	33	51	968
total	1603	155	137	22	61	101	2079

Table 6.11: Frequency of the progressive constructions per type of meaning of oblique markers of a temporal reference point.

The default in case no marker was found, was none. Table 6.10 shows the results.

A final aspect coded for was the presence of temporal reference points. From Lemmens' and Boogaart's descriptions, it seems that the ah-pattern is somewhat more abstract or less grounded than the cardinal posture verbs or, as Lemmens (2005, 211) states: "[The cardinal posture verb] construction is more typical in contexts where the action is viewed from a wider perspective, situated within the location at hand or as part of the setting. In contrast, the prep-progressive focuses on the action itself.". Another aspect of this grounding or contextualizing might be that for the four patterns derived from verbs denoting physical activities, more markers of temporal reference points are present. With markers of temporal reference points, I mean those oblique markers that denote at what point we should temporally locate the activity. These can be vague markers of time, such as 'lately' and 'a while ago', or precise markers of time, specifying a more precise time frame, such as 'friday last week', 'last year'. Other categories include deictic markers, relative markers and conditionals. Although all markers such as 'last week' are essentially deictic, with deictic I mean cases such as 'now' and 'at this moment'. relative markers are cases where the temporal setting of the activity is relative to another activity, such as 'when John came home' or 'after the game'. conditional markers, finally, mark not so much a temporal reference point, but rather a conceptual reference point, viz. a condition for the occurrence of the activity in the main clause. These conditions often coincide with a temporal reference point, but this need not be the case, e.g. in hypothetical and counterfactual conditionals. Table 6.11 shows the distribution. As we can see, the global prediction holds: more markers are relatively found with the four 'more concrete' patterns than with ah. In the remainder, this category will be called referencePoint.

Other determinants Two final determinants may be worth investigating. Firstly, the *Algemene Nederlandse Spraakkunst* (Haeseryn et al. 1997, 735) mentions that there may be register differences between the `ah`-construction and the other four, with the former being more formal. An explorative study on informal written data (UseNet conversations from the Condiv-corpus; Grondelaers et al. (2000)) as well as more formal spoken data (components *f* through *h* of the Corpus Gesproken Nederlands) showed that more formal spoken language indeed shows a more frequent `ah`-pattern, whereas the informal written language does not. Therefore, I chose to include the more formal spoken language as the formal counterpart of the informal spoken language of component *a* of the Corpus Gesproken Nederlands. Formal *written* language, finally, is problematic for a model such as AM, as it often comes heavily edited by multiple people.

The use of register raises an interesting question about the modeling enterprise. We assume, when doing a modeling study like the present, that the behavior of an individual does not deviate from that of the population. Because of this assumption, we can use data from different speakers to model the cognitive abilities of a single artificial agent. Obviously, this is an idealization, as all individuals behave also according to the norms of their subpopulations and language situations, which they must have stored as part of their linguistic competence. One aspect of this diversity, viz. formal versus informal language, can be easily used as a parameter in cognitive modeling studies such as the present one, so that the choice of a progressive is not only motivated by all kinds of structural, semantic and discours-pragmatic phenomena, but also by lectal properties.

This, in turn, raises the question whether we can also model regional differences, e.g. between Netherlandic and Belgian Dutch, with AM or any cognitive model. I think not, unless we are modelling the linguistic competence of a speaker fluent in both varieties. With AM being a model of one person's knowledge, we cannot assume that the input consists of equally many input data from one region as from the other. Different registers, on the other hand, *are* things many speakers are exposed to, so that it is not unlikely that these register-labeled exemplars are present in the memory of a single speaker. In general, Analogical Modeling is a perfect model to deal with register differences within one speaker. Differences between communities cannot be modeled with a cognitive model, and are therefore left out of this study.

Our formal register consists of the components *f*, *g* and *h*, which are made up of broadcasted interviews and discussion, political discussions and meetings and classroom situations. All of these situations call for a more formal register as the situations are institutionalized settings. These sections comprise slightly over a million words, against the 1.7 million of component *a*. Table 6.12 shows the difference between the formal section and the informal section. Strikingly, the relative number of progressives in general is much lower in the formal section than in the informal section.

Secondly, we can investigate in what types of clauses the progressives are used. The motivation for this factor is the idea that, as Lemmens stated, the cardinal posture verbs describe more of a contextualized situation. As such, clauses containing them can be thought to be heavier in terms of information and communicative value. Because of this, cardinal posture verb constructions and the `loop`-pattern are expected

	register		
construction	formal	informal	**total**
ah	168	585	753
lig	14	34	48
loop	9	81	90
sta	33	187	220
zit	138	830	968
total	362	1717	2079

Table 6.12: Frequency of the progressive constructions per register type.

	clause					
	adverbial	complement	main	question	relative	**total**
ah	87	62	512	54	38	753
lig	0	3	38	3	4	48
loop	7	4	70	0	9	90
sta	15	10	185	5	5	220
zit	53	41	833	27	14	968
total	162	120	1638	89	71	2079

Table 6.13: Frequency of the progressive constructions per clause type.

to occur more in main clauses than does the ah-progressive. The factor clause codes the different clause types: main for main clauses, question for polar and wh-questions, adverbial for temporal, conditional, causal etc. adverbial clauses, infinitive, for infinitive clauses, complement for complement clauses, and finally relative for relative clauses. In general, the global hypothesis that subordinate clauses attract the ah-construction holds, although sta occurs in a high number of adverbial, complement and relative clauses as well. For questions, furthermore, it is interesting to see that the ah pattern occurs twice as often as the zit-progressive. This is partly due to frequent questions such as *wat ben je aan het doen?* 'what are you doing', that are typically produced with an ah-construction.

6.3 Results

The datapoints were manually coded for the variables mentioned in the previous section and entered to the `raam` programme, using the a leave-one-out cross-validation regime. As in the previous experiments, the pointer statistics were used. For the following, the variables will be presented in order of (6.21).

(6.21) ⟨ `empathy agentivity verb domain aktionsart evaluative aspect location referencePoint aspectualMarking register clause`⟩

The tables show the F-scores at different ranges R_α. The graphs show the precision, recall and F-score for the five categories on the two most salient axes, namely those of increasing abstraction and decreasing specificity.

		Maximum α												
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Minimum α	**0**	0.24	0.51	0.64	0.66	0.66	0.67	0.67	0.67	0.67	0.66	0.66	0.66	0.66
	1		0.51	0.64	0.66	0.66	0.67	0.67	0.67	0.67	0.66	0.66	0.66	0.66
	2			0.64	0.66	0.66	0.67	0.67	0.67	0.67	0.66	0.66	0.66	0.66
	3				0.66	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67
	4					0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
	5						0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
	6							0.65	0.65	0.65	0.65	0.65	0.65	0.65
	7								0.65	0.64	0.64	0.64	0.64	0.64
	8									0.63	0.63	0.63	0.63	0.63
	9										0.59	0.59	0.59	0.59
	10											0.53	0.53	0.53
	11												0.27	0.27
	12													0.00

Table 6.14: F-scores for the `ah`-pattern at different ranges of abstraction.

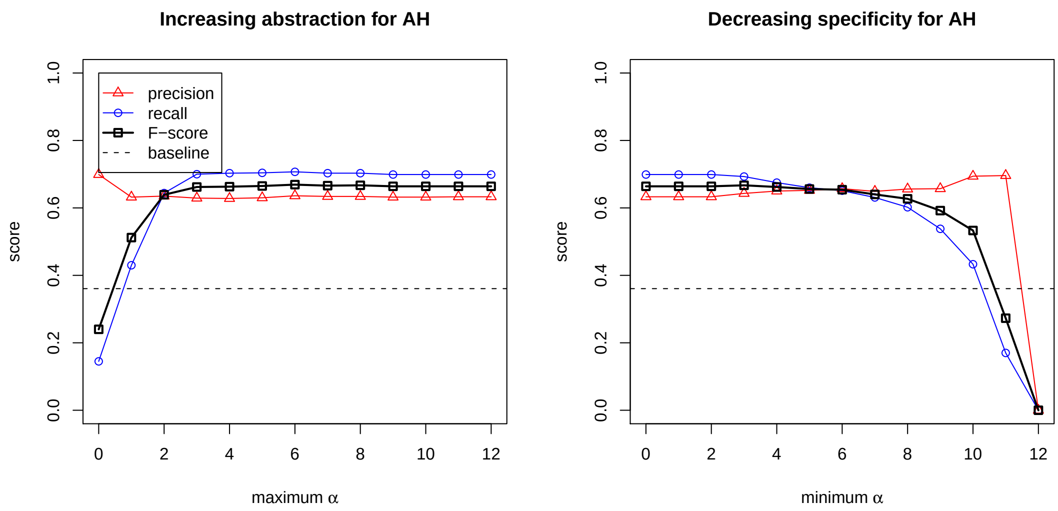


Figure 6.1: Scores for `ah` on the dimensions of increasing abstraction and decreasing specificity.

	Maximum α												
Minimum α	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
0	0.18	0.25	0.36	0.35	0.32	0.33	0.38	0.37	0.37	0.37	0.37	0.37	0.37
1		0.25	0.36	0.35	0.32	0.33	0.38	0.37	0.37	0.37	0.37	0.37	0.37
2			0.36	0.35	0.32	0.33	0.38	0.37	0.37	0.37	0.37	0.37	0.37
3				0.35	0.32	0.33	0.38	0.37	0.37	0.37	0.37	0.37	0.37
4					0.32	0.33	0.39	0.37	0.37	0.37	0.37	0.37	0.37
5						0.38	0.37	0.37	0.34	0.34	0.34	0.34	0.34
6							0.37	0.33	0.35	0.35	0.35	0.35	0.35
7								0.35	0.35	0.37	0.37	0.37	0.37
8									0.26	0.23	0.23	0.23	0.23
9										0.26	0.26	0.26	0.26
10											0.16	0.16	0.16
11												0.08	0.08
12													0.00

Table 6.15: F-scores for the `lig`-pattern at different ranges of abstraction.

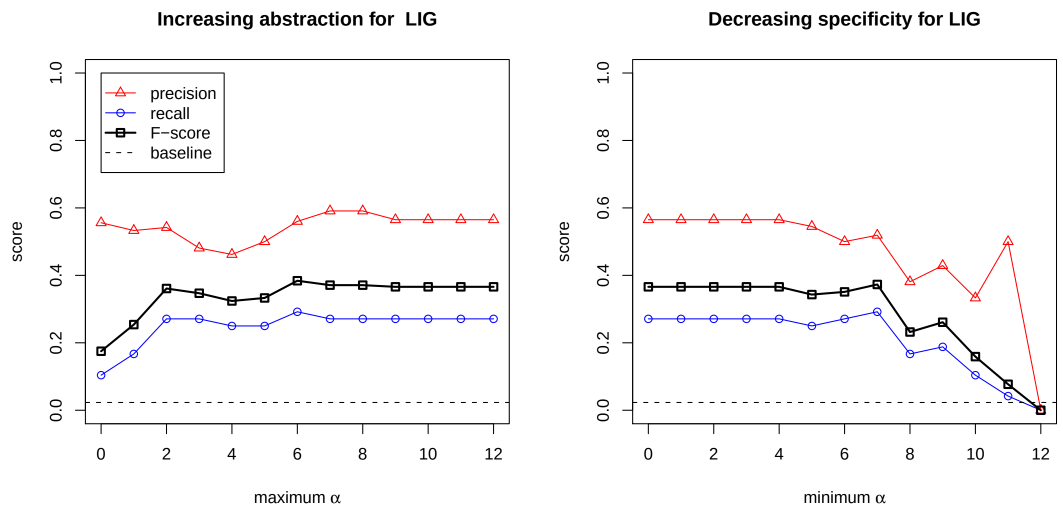


Figure 6.2: Scores for `lig` on the dimensions of increasing abstraction and decreasing specificity.

	Maximum α												
Minimum α	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
0	0.00	0.04	0.15	0.15	0.16	0.18	0.20	0.20	0.21	0.21	0.21	0.21	0.21
1		0.04	0.15	0.15	0.16	0.18	0.20	0.20	0.21	0.21	0.21	0.21	0.21
2			0.15	0.15	0.16	0.18	0.20	0.20	0.21	0.21	0.21	0.21	0.21
3				0.15	0.18	0.18	0.20	0.21	0.21	0.21	0.21	0.21	0.21
4					0.18	0.19	0.19	0.20	0.21	0.21	0.21	0.21	0.21
5						0.21	0.20	0.20	0.21	0.21	0.21	0.21	0.21
6							0.21	0.21	0.21	0.21	0.21	0.21	0.21
7								0.22	0.22	0.23	0.23	0.23	0.23
8									0.21	0.21	0.21	0.21	0.21
9										0.21	0.21	0.21	0.21
10											0.15	0.15	0.15
11												0.08	0.08
12													0.00

Table 6.16: F-scores for the `loop`-pattern at different ranges of abstraction.

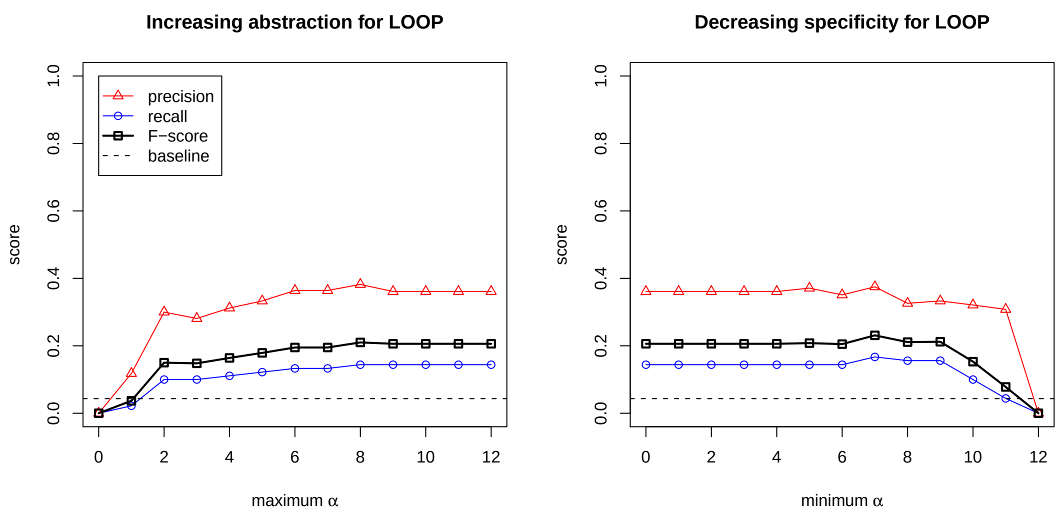


Figure 6.3: Scores for `loop` on the dimensions of increasing abstraction and decreasing specificity.

Minimum α		Maximum α												
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
	0	0.06	0.14	0.19	0.19	0.18	0.17	0.18	0.18	0.19	0.19	0.19	0.19	0.19
	1		0.14	0.19	0.19	0.18	0.17	0.18	0.18	0.19	0.19	0.19	0.19	0.19
	2			0.19	0.19	0.18	0.17	0.18	0.18	0.19	0.19	0.19	0.19	0.19
	3				0.20	0.19	0.18	0.19	0.19	0.19	0.19	0.19	0.19	0.19
	4					0.19	0.19	0.19	0.20	0.20	0.20	0.20	0.20	0.20
	5						0.18	0.20	0.19	0.19	0.19	0.19	0.19	0.19
	6							0.19	0.19	0.20	0.20	0.21	0.21	0.21
	7								0.21	0.21	0.22	0.21	0.21	0.21
	8									0.21	0.21	0.21	0.21	0.21
	9										0.21	0.21	0.21	0.21
	10											0.20	0.20	0.20
	11												0.12	0.12
	12													0.00

Table 6.17: F-scores for the `sta`-pattern at different ranges of abstraction.

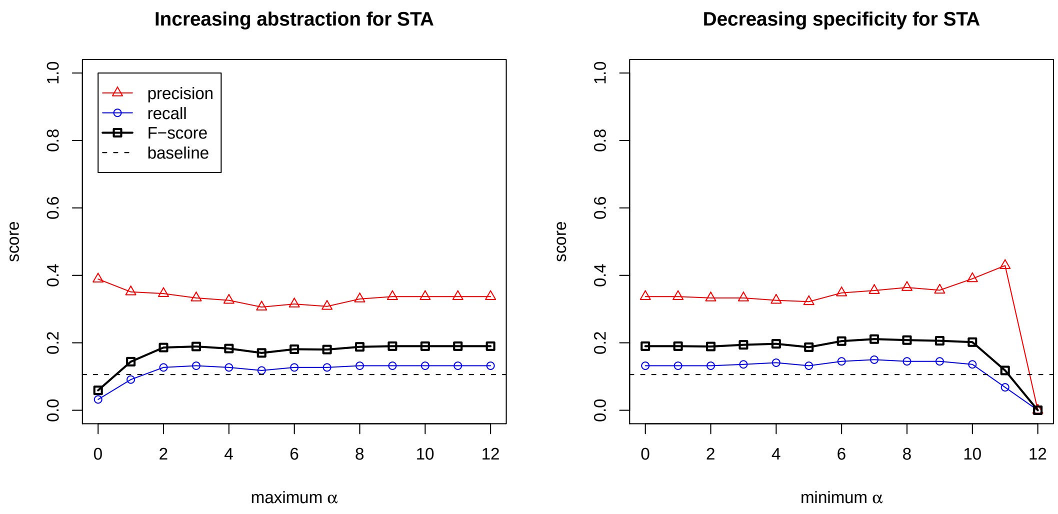


Figure 6.4: Scores for `sta` on the dimensions of increasing abstraction and decreasing specificity.

		Maximum α												
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Minimum α	**0**	0.42	0.59	0.64	0.66	0.67	0.67	0.67	0.66	0.67	0.67	0.67	0.67	0.67
	1		0.59	0.64	0.66	0.67	0.67	0.67	0.66	0.67	0.67	0.67	0.67	0.67
	2			0.64	0.66	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67
	3				0.66	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67
	4					0.67	0.67	0.67	0.66	0.67	0.67	0.67	0.67	0.67
	5						0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65
	6							0.59	0.59	0.59	0.59	0.59	0.59	0.59
	7								0.57	0.57	0.57	0.57	0.57	0.57
	8									0.55	0.54	0.54	0.54	0.54
	9										0.51	0.51	0.51	0.51
	10											0.42	0.42	0.42
	11												0.14	0.14
	12													0.00

Table 6.18: F-scores for the `zit`-pattern at different ranges of abstraction.

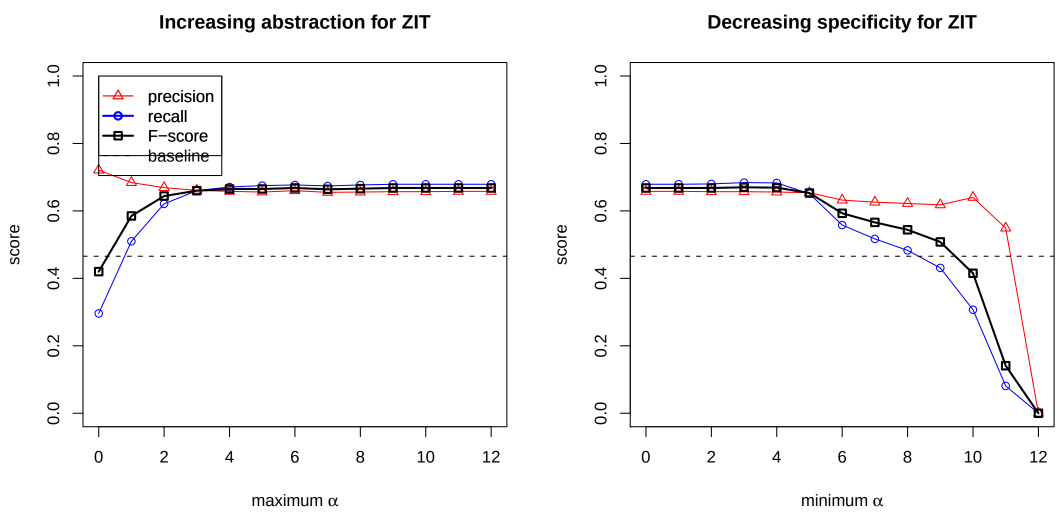


Figure 6.5: Scores for `zit` on the dimensions of increasing abstraction and decreasing specificity.

6.4 Discussion and qualitative exploration

6.4.1 Global interpretation

The two biggest categories, `zit` and `ah`, are predicted reasonably well, with the highest F-score for both categories lying around $F = 0.67$. This means that we can explain two-thirds of the variation for these categories. Although the factors eventually driving the alternation have clearly not been found yet, the results for these two categories are promising. Moving towards the analysis of increasing abstraction and decreasing specificity, we can see that `ah` reaches its asymptotic high at $R_\alpha = [0 \ldots 6]$, although the difference with the model at $R_\alpha = [0 \ldots 3]$ is less than one percent ($d = 0.007$). We could say that with supracontexts leaving at most three slots out of twelve open, the model is reasonably able to generalize towards unseen cases. On the side of reducing specificity, the right panel of figure 6.1, we can see that the model starts deterioration at $R_\alpha = [6 \ldots 12]$ and in the models allowing for less specificity. The precision scores (red triangles) in this plot show us that this drop in accuracy is not caused by too much overgeneralization, but rather by an increasing amount of undergeneralization, as can be seen in the recall scores (blue circles). This means that for many novel cases there are either other categories overruling the abstract schemas pointing towards an `ah`-outcome, or there are no reliable, i.e. homogeneous supracontexts in the data.

For `zit`, the story is slightly different. Increasing the allowed amount of abstraction quickly ceases to improve the F-score of the model. At $R_\alpha = [0 \ldots 3]$, the model is already close to its asymptote, which is eventually reached at $R_\alpha = [0 \ldots 6]$. Again, little abstraction is necessary in predicting the `zit`-progressive. Decreasing specificity has a stronger effect. The precision decreases slowly, as it did in the `ah`-progressive, but the recall drops rapidly after $R_\alpha = [4 \ldots 12]$. This means that if we exclude supracontexts containing four or less open slots, the model predicts thirty-one more observed cases of `zit` wrongfully: the number of true positives for `zit` at $R_\alpha = [4 \ldots 12]$ is 661, against 630 at $R_\alpha = [5 \ldots 12]$. It seems that this construction, unlike the `ah`-progressive has an insufficiently coherent core to be predicted on the basis of more abstract supracontexts alone. In Wittgensteinian terms, the `zit`-progressive is a construction showing family resemblances.

The pattern of `sta` is predicted rather poorly. With a baseline of 0.106, i.e. the relative frequency of the pattern, its top F-score lies only some 10% higher, namely at $F = 0.217$, for a model allowing only relatively abstract supracontexts ($R_\alpha = [7 \ldots 9]$). The major problem for the model is to find coherent supracontexts that allow for generalization; the low recall scores (blue circles) in the two plots of the two salient dimensions of abstraction display this (figure 6.4). This means that given the set of determinants the model fails to generalize towards observed cases of the `sta`-progressive. Perhaps this is due to the fact that the lexical meaning of *staan* is more present than in *zitten*, but the codable symptoms of this lexical meaning (e.g. more obliques of time and place) are absent, so that the model will often use the more widely used and relatively bleached pattern of `zit` instead. A glance at the confusion matrix of the model at $R_\alpha = [7 \dots 9]$, the best model for the `sta`-progressive, learns us that this is indeed the case. Out of the 186 false negatives, i.e. observed cases of `sta` predicted to be otherwise, more than half (95) are predicted to be *zitten*. Only in 51 cases `ah` is falsely selected, whereas the `ah`-progressive is much more than half as frequent as the `zit`-pattern, so that this generalization of `zit` towards `sta` is not merely due to the high relative frequency of `zit`.

Given this poor performance, the curves in the plots display the 'regular' pattern. The asymptote is quickly reached when we add layers of abstraction, so that we can claim that with relatively little abstraction, the model is already at the top of its potential. When decreasing the minimum α, the model rises 0.021 in F-score, from $F = 0.190$ at $R_\alpha = [0 \dots 12]$ to $F = 0.211$ at $R_\alpha = [7 \dots 12]$. Whether we should see any significance in this pattern, is not directly obvious.

Fourthly, the `lig`-progressive is predicted relatively well, despite its low relative frequency, and hence baseline performance of 0.023. It seems that the `lig`-progressive has a clear profile, in terms of the sets of values it surrounds itself with (locational markers denoting a spatial relation to an object, verbs denoting passive actions). At the peak ($R_\alpha = [4 \dots 6]$), the F-score is $F = 0.394$, more than seventeen times the baseline model. The peak in the left panel of figure 6.2 has a maximum of $\alpha = 6$ as well. Allowing more abstract supracontexts does not further improve the model, and even hurts the performance, albeit minimally.

Reducing the allowed concrete supracontexts has influence from a minimum $\alpha = 8$ onwards. Using patterns with maximally five values out of a possible twelve specified thus does not harm the model. When the maximum number of allowed values is set to four (i.e. $\alpha \geq 8$), the performance deteriorates.

`loop`, finally, also performs reasonably well given its baseline performance model which is based on the fact that only 4.3% of all progressives are cases of the `loop`-pattern. At its peak, the model performs almost five times better, with an F-score of $F = 0.231$ (at $R_\alpha = [7 \dots 9]$, $R_\alpha = [7 \dots 9]$, $R_\alpha = [7 \dots 9]$, and $R_\alpha = [7 \dots 9]$). Three things are remarkable. Unlike in most cases we have seen so far, the model only gradually approaches its asymptotic high when we increase the allowed abstraction level (left panel of figure 6.3). The peak of this curve is reached only at $R_\alpha = [0 \dots 8]$. Secondly, unlike in all other cases, for the `loop`-progressive, the model that allows for maximally one open slot in the supracontextual schemas, viz. $R_\alpha = [0 \dots 1]$, performs under the baseline. Finally, in increasing the minimum level of α, we see that the model improves slightly, reaching a peak at $R_\alpha = [7 \dots 12]$. A model with less supracontexts with many features specified for this category leads to a better performance.

These three phenomena point, in my opinion, to the fact that the `loop`-progressive prefers abstraction over concreteness. This, to my mind, means that the `loop`-pattern does not have a series of related, more specific patterns, but rather a smaller set of more abstract patterns, as we will see in section 6.4.3. It is not a typical family-resemblance category, but rather a pattern driven by a few, synchronously hardly overlapping meanings, namely (analytically speaking) 'moving around doing something', 'doing something repetitively (and hence to someone's annoyance)'.

Globally speaking we can say that the patterns are reasonably well predicted, except for the `sta`-progressive. In increasing the allowed level of abstraction (more open slots), the models approach the peak in their performance relatively early, except for `loop`, which remains climbing up to $R_\alpha = [0 \ldots 8]$. The deterioration patterns for the decrease in specificity differ, with `sta` dropping steeply only at $R_\alpha = [10 \ldots 12]$, `zit` already at $R_\alpha = [6 \ldots 12]$, and the other three categories somewhere in between. It is hard to make any general statements about the level of abstraction at which this family of constructions is most likely stored. The `loop`-progressive seems to favor a lack of specificity, but such a lack would deteriorate the performance for the `zit`-pattern. Except for the `loop`-progressive, most categories reach their asymptote at a very low level of allowed abstraction. The fact that the model predicts the `loop`-pattern on low levels of abstraction ($R_\alpha = [0 \ldots 2]$, $F = 0.150$) about 71% as good as when it allows more abstraction ($R_\alpha = [0 \ldots 8]$, $F = 0.210$) while the other patterns remain almost at the same F-score, suggests that for a fuller empirical coverage of the data, a model with slightly more abstraction allowed (e.g. $R_\alpha = [0 \ldots 8]$) is optimal. Perhaps it is the safest thing to say that some abstraction is necessary, but that we can say little about whether concrete supracontexts should be used.

6.4.2 An example

To see some of the categorization dynamics at work in changing the R_α, let us focus on the following example and its given context.

(6.22) *toch niet die Jeroen die jou half aan 't verkrachten was toen in de*
PRT not that Jeroen that you half on the rape.INF were then in the
woonkamer ?
living.room ?

'It wasn't that Jeroen guy, who was more or less raping you, back then in the livingroom ? **(fn000546.303)**

(6.23) ⟨`human agent verkrachten stativeMotion activity no onvoltooid space vague none informal relative`⟩

Table 6.19 shows the supracontextual scores per level of abstraction, whereas table 6.20 shows the familiar triangle of minimum and maximum levels of α, but now filled with the predictions for our new exemplar. Showing all 244 supracontexts used, would be too page-consuming. The development of the predictions is quite interesting for this case. If we focus on the axis of increasing the abstraction, we see that the first prediction can be made at $R_\alpha = [0 \ldots 2]$, that is: using a schemas that have only two open slots. For this setting, the erroneous prediction `sta` is made. This remains the progressive construction of choice until $R_\alpha = [0 \ldots 6]$, when the correct

	agreements per outcome				
α	`ah`	`lig`	`loop`	`sta`	`zit`
2	0	0	0	1	0
3	1	0	0	8	1
4	15	0	3	26	1
5	56	0	3	43	6
6	72	0	1	38	14
7	31	0	1	26	16
8	1	0	0	3	9
9	0	0	0	0	2

Table 6.19: Supracontextual scores for the exemplar in (6.23).

`ah`-construction is predicted. This means that in this case, increasing the allowed abstraction in the schemas used to predict the item has a beneficial effect: the 'noise' of the `sta`-schemas is overruled by supracontexts predicting the correct `ah`-progressive.

Decreasing the specificity has little effect up to $R_\alpha = [7 \ldots 12]$. After that point, the category that only marginally played a role, viz. the `zit`-pattern becomes dominant, and is in fact the only construction yielding homogeneous supracontexts at the very abstract level of $\alpha = 9$, as can be seen in table 6.19. The succesful prediction of `ah` thus relies on supracontexts in a medium range of abstraction. If concrete supracontexts that are too concrete are used, the `sta` pattern is dominant; if only the highly abstract ones can be employed, `ah` does not contribute any homogeneous schemas anymore and the category of `zit`, which is has relatively little influence on lower ranges of abstraction, produces the best predictor.

It is important to remember that these predictions are eventually grounded in the exemplars over which the supracontextual analysis is executed. In fact, the 244 supracontexts are derived from a set of only eight exemplars. In the style of the Nearest Neighbor algorithm, we can order these exemplars according to their distance to the new item. We calculate the distance by counting the number of mismatching values between the two items. Figure 6.6 does so by given the vectors of values in boxes. The number to the left gives the distance to the new item. The four nearest neighbors are given in examples (6.24) through (6.27).

(6.24) *toen stond ze op de kant dubbel zo op en neer te gaan .*
then stood she on the shore double like.that up and down to go.INF .

'Then she was bouncing up and down twice as (. . .) on the shore.
(fn000555.144); distance = 2

(6.25) *daar hoorde ik degene die daar aan het vangen was ...*
there heard I the.one who there on the catch.INF was ...

'There I heard the person who was catching there ...'
(fn007494.124); distance = 3

(6.26) *die ook uh gewoon ergens staan te grazen ...*
that also uhm just somewhere stand to graze.INF ...

'that are also just grazing somewhere ...'
(fn000830.162); distance = 3

(6.27) *heeft ie toch altijd zo'n pennetje in z'n hand waar ie mee loopt te draaien ...*
has he PRT always such.a pencil.DIM in his hand where he with walks to turn ...

'And then he's always holding a little pencil in his hand with which he's twirling ...'
(fn000981.176); distance = 3

The structured representation in figure 6.6 reveals much of the supracontextual predictions. Firstly, `sta` is predicted for the ranges with little abstraction. This is because the 'nearest neighbor' is an instance of the `sta`-progressive. As such, this exemplar underlies many supracontextual schemas that fit the new item. Its effect is reinforced by a highly similar exemplar, namely the one based on utterance (6.26), which differs one more value from the new item (viz. the `empathy`).

The representation also shows how three exemplars in the `ah`-progressive are closely related. At a distance of 3, the exemplar based on utterance 6.25 has a different `verb`, `referencePoint` and `register` than the new item. Together with two more exemplars that also differ in the `domain`, this exemplar forms the basis of many supracontexts that weigh heavily due to the fact that the squared or agreement measure is used. On even higher levels of abstraction ($\alpha \geq 5$), the gang of three exemplars forms supracontexts based on themselves and the bottom right exemplar as well, thus making even stronger predictions (a supracontext with four exemplars underlying it yields sixteen agreements). This gang-effect is important, as it shows that the nearest neighbor might not always be the correct predictor, but that it may be outweighted by a cluster of more distant exemplars.

Finally, an exemplar predicting `zit` is found at a distance of 4. At the very high levels of abstraction ($R_\alpha = [8 \dots 12]$ and ranges with slightly more minimum abstraction), this exemplar is the only one contributing supracontextual schemas that match the new item and are homogeneous. This is interesting, as there are three more exemplars at the same distance predicting `ah`. Apparently, given our dataset there are schemas based on the `zit`-exemplar that are homogeneous at this abstract level of comparison. The interesting ones, at the highly abstract levels ($\alpha \geq 8$), are cases in

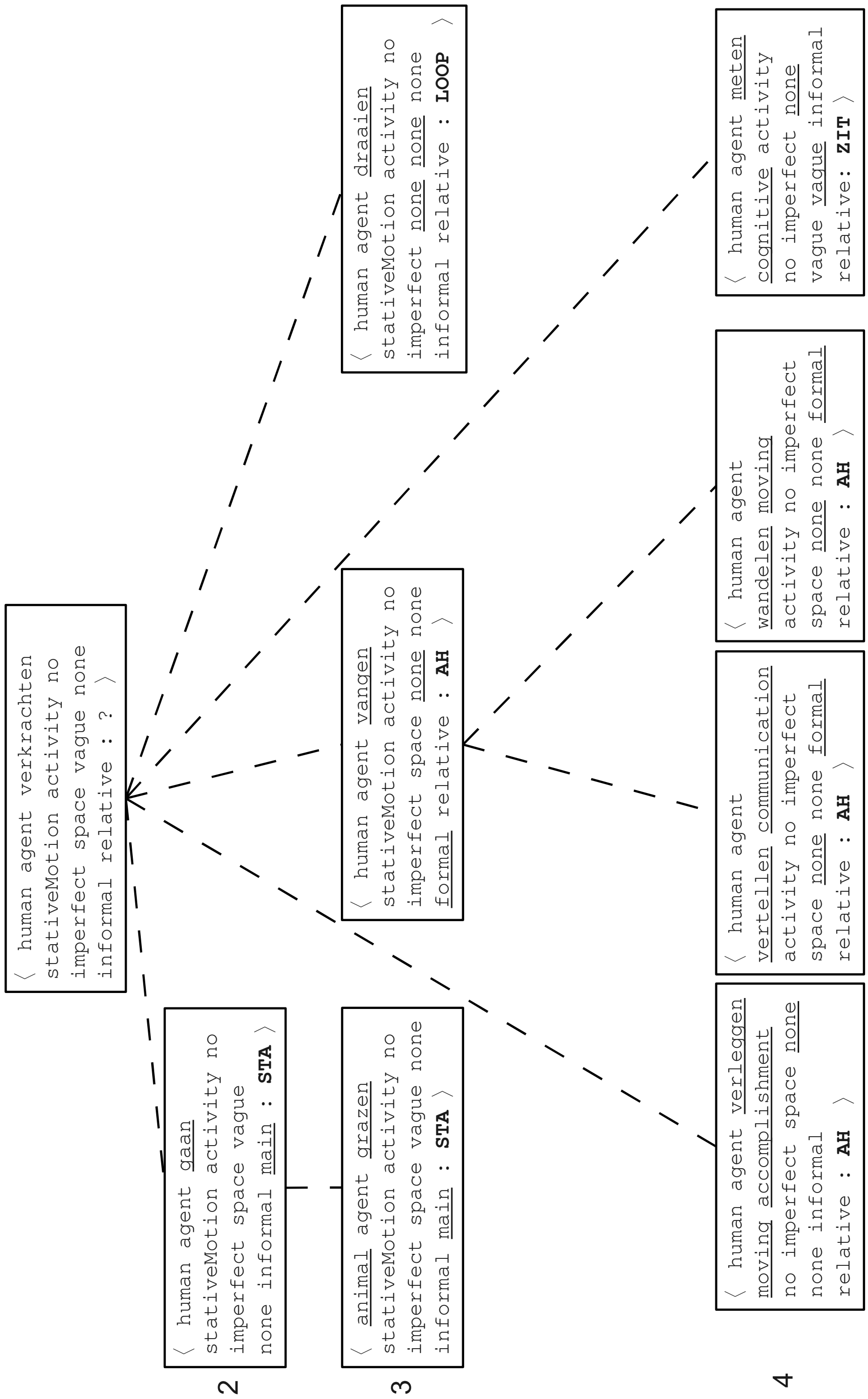


Figure 6.6: The ‘nearest neighbors’ of the exemplar in (6.23)

	Maximum α												
Minimum α	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
0	TIE	TIE	sta	sta	sta	sta	ah	ah	ah	ah	ah	ah	ah
1		TIE	sta	sta	sta	sta	ah	ah	ah	ah	ah	ah	ah
2			sta	sta	sta	sta	ah	ah	ah	ah	ah	ah	ah
3				sta	sta	sta	ah	ah	ah	ah	ah	ah	ah
4					sta	ah	ah	ah	ah	ah	ah	ah	ah
5						ah	ah	ah	ah	ah	ah	ah	ah
6							ah	ah	ah	ah	ah	ah	ah
7								ah	ah	ah	ah	ah	ah
8									zit	zit	zit	zit	zit
9										zit	zit	zit	zit
10											TIE	TIE	TIE
11												TIE	TIE
12													TIE

Table 6.20: Predictions for the exemplar in (6.23).

which it is specified that the `clauseType` is `relative`, that the `referencePoint = vague` and in which one or two out of the following properties are present:

- `empathy = human`,
- `agentivity = agent`,
- `domain = activity`,
- `evaluative = no`,
- `aspect = imperfect`,
- `register = informal`

With these highly abstract, but nevertheless homogeneous schemas, the category of `zit` is the strongest in the more abstract regions of the supracontextual comparisons.

6.4.3 The descriptive value of AM

Thus far, the discussion of the results has been very 'cognitive', except perhaps for the analysis of the significance of the most frequently used supracontexts of the dative constructions in chapter 4. We have been working on the question of abstraction, and several interesting results were found in the analysis of the nine categories so far. This section is devoted to the question how we can use the insights from Analogical

Modeling for language description, for instance for second language learners or in writing reference grammars. More specificly, suppose we would want to explain how to use one of the progressive constructions. In that case, stating hard rules such as 'if the verb expresses a speaker's negative judgement, use `loop`' would be useless, as the data is too variable and as probabilistic rules are not fit for explanatory purposes.[2] Such a description has to be true to the spirit of AM. We are expected to read in a description the association between a single variable and the category labels. This global significance of a variable may be interesting, but we cannot say much about it from the results of an analysis using AM. Contingency tables, as given throughout this study, do give such information, and should be consulted if we want global information on variables. In this section, other means of providing information on the grammar than global associations are explored.

Furthermore, we cannot simply list all the exemplars we found in some corpus either, as this yields no description, but mere enumeration of examples. How do we recognize the importance of concrete cases in descriptions without losing the generalization that rules provide us with explanations or grammatical descriptions of a category? I believe this can be achieved through the presentation of a small set of prototypical examples, enriched with analytical judgements about overlap between them. In chapter 4, we have been looking at the prototypical supracontexts, which also have some explanatory value, but now we will be looking at the prototypical exemplars. Both are means of providing descriptive information on the function of the patterns, without resorting to rules and global associations. This section thus is an aside that leaves the issues of abstraction and schematization for what they are, and focusses on another aspect of linguistics. I believe this is licenced amongst others by the fact that this is the first study dealing with functional approaches to grammar from an AM perspective, and it seems valuable to show its potential beyond being a cognitive model.

How do we determine how prototypical an exemplar is? I believe the prototypicality can be estimated by measuring how often that exemplar is used as an analogue for other cases. Now, if an exemplar is very similar to other cases, it will underly many supracontextual schemas and it can thus be said to be more influential than distant items. We therefore do not count the *presence* of an exemplar as an analogue for a new item, but rather the number of supracontextual schemas it supports per predicted new item. Thus if for a new case some exemplar #422 is found in five homogeneous supracontexts that predict that new case, five points are added to the total score of this exemplar. If we do so for all members of a category we are trying to predict, for instance the `loop`-progressive, we get a profile of the exemplars that are the most important in predicting members of that category. Obviously this profile also includes exemplars of other categories that are often used in predicting the `loop`-progressive, but for now let us focus only on the cases where exemplars of the `loop` category

[2] A statement like 'if the verb expresses a speaker's negative judgement, use `loop` in 65% of the cases' does not help a second language learner much either.

predict other cases of the `loop` category.

Table 6.22 displays the twenty most frequently used exempars. Let us look at the description of the `loop`-progressive this set can give use. First of all, a very salient cluster is that of verbs of communication, and then especially those that denote some sort of complaining. These activities are furthermore often negatively evaluated by the speaker.[3] Some extensions of this prototypical set of exemplars can be found in verbs and idioms expressing being overly critical in an unfriendly way to someone (*afzeiken*, lit. 'off.piss'), sucking up (*slijmen* lit: 'slime'), having a fight (*ruzie maken*, coded as `maken`, lit: 'make fight'), communicating by making much noise (*roepen*, 'yell') and non-communicative emissions of sound, such as *steunen*, 'groan'). This set of apparently similar cases can be said to form one usage type of this construction.

A second one is closely related and some exemplars seem in between the two sets.[4] For instance, the verbal idiom *een rel schoppen* 'start a riot'; lit: 'kick a riot' can be placed with the previous cluster, as it is a social, possibly communicative activity that bears the element of 'negative communication' or 'face-threatening communication' found in many members of the previously described set. It also has a telic Aktionsart, because there is a clear result state from this process, namely that there is a riot.

As such, it belongs to a cluster of other telic verbs (accomplishments and achievents). This is what I would like to call the second usage type of the `loop`-progressive. They are marked in the table with an 'X' under the rightmost column. Strikingly, in all of these cases, the grammatical aspect is also perfective. The interaction between the progressive, the perfective aspect and the verbal telicity in these cases is rather interesting. Consider the sentences (6.28)-(6.29. Both express a process that is telic, but the former has a duration, whereas the latter typically does not. In interaction with the `loop`-progressive (as well as the perfective grammatical aspect), now, we can see that the `loop`-progressive does not function merely as an indicator of the fact that an event has been ongoing, but also as a way of intensifying the action somehow. Changing the progressive to `zit`, which is perfectly possible, enables a different set of inferences, to my intuition. With `loop`, it is more likely that the action is uninterrupted than with `zit` and it somehow seems that there is, again, an element of the evaluative or empathic involvement of either the speaker or the referent of the grammatical subject. If in sentence (6.29) they have been causing a fight, with the `loop`-construction it seems that 'they' care very much about the topic they have been fighting about, whereas this seems to be less so with `zit`.

[3]Interestingly enough, the simplest verb to denote the activity of complaining, *klagen*, often expesses a negative evaluation on the speakers part as well. Perhaps this has led to the emergence of periphrastic forms in which the same verb root is used nominally in all sorts of periphrastic constructions, such as *je beklag doen* 'make a complaint' and *een klacht indienen* 'make a complaint'.

[4]Which is exactly the reason why one would not want to have hard-coded prototypes.

(6.28) *uhm ja ik heb er zelf ook een hele tijd naar lopen zoeken*
ehm yes I have there self also a whole time to walk.INF search.INF
naar videorecorders .
to video.recorders .

'Ehm yeah, I have been looking there for video recorders quite a while too.' **(fn000318.376)**

(6.29) *en die hebben daar een ontzettende rel over lopen schoppen .*
and those have there a enormous riot about walk.INF kick.INF .

'And they have been causing a huge fight over that.' **(fn000816.100)**

The other values do not reveal much more, except maybe for (1) the fact that the most typical exemplars are from the informal part of the CGN and (2) that in six out of twenty cases a vague marker of duration was present, which is consonant with the idea that loop also intensifies the action in some way. Finally, it is remarkable that no cases of loop with an actual verb that denotes an activity involving motion, were found among the best cases. From these results, it seems that this usage type of loop-progressive is not a salient, extendable prototype. Follow-up studies will have to assess this hypothesis further.

A description of the loop-progressive on the basis of the most prototypical exemplars would look like the following.

> The loop progressive is used with verbs denoting communication that is typically seen as negative, such as *complaining* or *verbally abusing someone*. The second domain of verbs is that of telic verbs. This group is often used in the perfective form. In both groups, the subject of the clause is typically an agent that is affectively or empathically engaged in the action.

Secondly, let us look at the two bigger patterns and how the most salient exemplars can help us differentiate between them. Table 6.23 and 6.24 give the twenty most often used exemplars per progressive type.

In the types of subjects, we can already see strong differences. Whereas the ah-pattern is used most often with abstract undergoing subjects, human beings (including speakers and hearers), that are conceptualized as experiencing or agentive, are the typical subjects of the zit pattern. Note here that this does not mean that these prototypical cases are the majority values per category. There are only 36 abstract subjects for the ah-progressive, against 271 speakers and 127 hearers. Nevertheless, empathy = abstract has the strongest association with the pattern, and associations between one category and a value are not only shaped by frequency, but also by dissociation of that value and other patterns.

Typical domains for the zit progressive include cognition ('think'), ingestion ('eat') and evaluation ('be fed up'). For the ah-progressive, the typical domains include cultural processes ('society changing', 'the language becoming more and more

	progressive	
	ah	zit
subjects	abstract, undergoer	human agent/experiencer
domains	cultural, stative physical	cognition, ingestion, evaluation
Aktionsart	activities, gradual completion	states, activities
aspect	imperfect	any
obliques	few	many
register	formal	informal
clause type	any	main clauses

Table 6.21: A description of the ah and zit-progressives.

influenced by English') and stative physical processes, such as 'growing' and 'become old'. All twenty typical ah-progressives are used in the imperfect tense, whereas the zit progressive displays two perfective and one ingressive tense.

Among the best cases of the ah-progressive, few spatial, temporal, and aspectual markers are found. This number is much higher for the zit-progressive, which corresponds to Lemmens's (2005) statement that the cardinal posture verb progressives are more 'situated' than the ah-pattern.

For the register, we can recognize what was already shown in the first exploration of the corpus. The zit-pattern is a more informal pattern, and only two out of the best twenty cases were from the formal part of the corpus. For the ah-pattern, this number is much higher, viz. eleven.

Finally, the clause types differ. Whereas the prototypical zit-progressive is used in main clauses, the ah-pattern is found across a wider number of constructions. This fact may also be explained by Lemmens' observation that the zit-pattern is more situated than the ah-pattern. Being more situated then means that the speaker takes more effort in elaborating the situation, with different obliques, but also by placing it on the foreground of the communication, and hence in the main clause rather than in a relative or adverbial clause, which function more as backgrounds.

Summarizing, the difference between the two major categories is multivariately motivated along the lines displayed in table 6.21.

6.5 Conclusion

In this experiment, I modeled a categorization task with five outcomes that is motivated by many distinct variables, so as to gain more insight in the effect of increasing abstraction and reducing specificity. Furthermore, this corpus experiment constitutes the first systematic modeling experiment of the Dutch progressive on the basis of spoken language material. The descriptive value of this work can be shown with the same means of the AM-analysis itself. By looking at the exemplars that most often underly supracontexts that predict items of one category, we can get a good profile of the typical cases.

In reducing specificity and increasing abstraction, most of the categories from this pattern behaved the same as three of the cases we saw before. In these cases, the increase of abstraction quickly rises and reaches its asymptote with very concrete patterns. The fact that this pattern occurred in the `NP`-dative, relative clause interpretations as well as the `zit` and `ah`-progressives, suggests that this is the normal behavior of any highly-frequent category, as all of these categories made up more than 40% of the cases. For these major patterns, the decrease of specificity also shows that these categories can be classified without remembering much detail. It seems that if we have the right determinants, highly-frequent patterns can be predicted without abstraction or without specificity. As such, they do only provide proofs of existence of both models, but no comparison.

More of a comparison can be found among the smaller categories. We have seen how the `PP`-dative displays a peak in its performance at a low level of maximum abstraction, and lower scores with more abstraction. The `loop`-progressive behaves radically different. Here, the F-score continues to rise if we increase the abstraction. Furthermore, and unlike all other categories, it also continues to improve if we decrease the amount of specificity, up to a peak at $R_\alpha = [7 \ldots 12]$. This might mean two things. The first interpretation is that the two deviating patterns of the `PP`-dative (little abstraction) and the `loop`-progressive (no early asymptote, continuing improvement) are accidental and the best range for all cases is by simply allowing all levels of abstraction ($R_\alpha = [0 \ldots n]$, where n is the number of variables).

A more interesting hypothesis is that for some sets of constructions (e.g. all progressive constructions in Dutch), the schemas are stored up to a more abstract level than for other constructions. This might mean that for general purpose aspectual patterns, more abstract and less specific constructions are stored than for argument structure patterns, for instance. If this is the case, language users must somehow keep track of an optimal range of supracontextual schemas. Langacker's (2009) idea that all schemas are inherent to a processed exemplar is consonant with this view. Given that all schemas are present, the model predicts that per constructional choice a certain bias towards a certain range of schemas is present.

n	exemplar	comm.	compl.	telic
976	⟨human agent mopperen communication activity yes imperfect none none none formal relative ⟩	X	X	
960	⟨hearer agent slijmen communication activity yes imperfect none none none informal adverbial ⟩	X		
864	⟨human agent zeiken communication activity yes ingressive none none none informal main ⟩	X	X	
736	⟨speaker agent zoeken generic accomplishment no perfect none none vague informal main ⟩			X
640	⟨human agent mopperen communication activity yes imperfect none none none informal complement ⟩	X	X	
640	⟨speaker agent slijmen communication activity yes imperfect none none none informal complement ⟩	X	X	
608	⟨speaker agent zoeken generic accomplishment no perfect none none none informal main ⟩			X
576	⟨human agent klagen communication activity yes imperfect none none vague informal main ⟩	X	X	
576	⟨human agent klagen communication activity yes imperfect none none vague informal main ⟩	X	X	
496	⟨human agent zeiken communication activity yes ingressive none conditional none informal main ⟩	X	X	
480	⟨human agent starten generic achievement no perfect none none none informal main ⟩			X
472	⟨human agent plakken stativeMotion achievement no perfect none none none informal main ⟩			X
464	⟨human agent mieren communication activity yes imperfect none none none formal relative ⟩	X	X	
448	⟨hearer agent afzeiken communication activity yes imperfect none none none informal adverbial ⟩	X		
416	⟨human agent schoppen social achievement no perfect none none none informal main ⟩			
416	⟨human agent maken communication activity no perfect none vague none informal main ⟩	X		
388	⟨human agent roepen communication activity no imperfect none none none informal main ⟩	X		
385	⟨human agent steunen emission activity no imperfect none conditional none informal main ⟩			
384	⟨human agent steunen emission activity no imperfect none none vague informal main ⟩			
384	⟨human agent roepen communication activity no imperfect none none vague informal main ⟩	X		

Table 6.22: The twenty most frequently used exemplars of the `loop`-progressive.

n	exemplar
11584	⟨speaker experiencer denken cognitive state no onvoltooid none vague none informal main ⟩
10820	⟨speaker agent afvragen cognitive state no onvoltooid none none none informal main ⟩
10768	⟨speaker experiencer denken cognitive state no onvoltooid none deictic none informal main ⟩
10768	⟨speaker experiencer denken cognitive state no onvoltooid none deictic none informal main ⟩
10428	⟨speaker agent afvragen cognitive state no onvoltooid none vague none informal main ⟩
9188	⟨speaker experiencer denken cognitive state no onvoltooid none vague none formal main ⟩
7884	⟨speaker experiencer denken cognitive state no onvoltooid none vague none informal complement ⟩
7810	⟨hearer agent eten ingestion activity no onvoltooid none none none informal main ⟩
7544	⟨human subject wachten stativePhysical state no onvoltooid none none none formal complement ⟩
7488	⟨human experiencer denken cognitive state no onvoltooid none vague none informal main ⟩
7280	⟨speaker experiencer balen evaluation state no onvoltooid none none none informal main ⟩
6432	⟨speaker experiencer balen evaluation state no onvoltooid none vague none informal main ⟩
6208	⟨human experiencer balen evaluation state no onvoltooid none none none informal main ⟩
6063	⟨hearer agent eten ingestion activity no onvoltooid space none none informal main ⟩
5968	⟨speaker experiencer denken cognitive state no onvoltooid none deictic vague informal main ⟩
5923	⟨hearer agent eten ingestion activity no onvoltooid none none none informal adverbial ⟩
5904	⟨speaker experiencer denken cognitive state no onvoltooid none deictic repetitive informal main ⟩
5874	⟨speaker agent drinken ingestion activity no voltooid space precise none informal main ⟩
5816	⟨speaker experiencer nadenken cognitive activity no voltooid none none none informal main ⟩
5552	⟨hearer agent eten ingestion activity no ingressief none none none informal main ⟩

Table 6.23: The twenty most frequently used exemplars of the `zit`-progressive.

n	exemplar
17536	⟨abstract subject worden cultural gradualCompletion no onvoltooid none none none informal main ⟩
16088	⟨abstract subject worden cultural gradualCompletion no onvoltooid none none none formal main ⟩
9744	⟨abstract subject veranderen cultural gradualCompletion no onvoltooid none none none informal main ⟩
9624	⟨speaker subject worden cultural gradualCompletion no onvoltooid none none none formal main ⟩
9616	⟨abstract subject verengelsen cultural gradualCompletion no onvoltooid none none none informal main ⟩
9592	⟨abstract subject vergrijzen cultural gradualCompletion no onvoltooid none none none formal complement ⟩
9080	⟨abstract subject worden stativePhysical gradualCompletion no onvoltooid none none none informal main ⟩
8424	⟨abstract subject worden evaluation gradualCompletion no onvoltooid none none none formal complement ⟩
7692	⟨abstract subject zakken generic gradualCompletion no onvoltooid none none none formal complement ⟩
7648	⟨speaker agent rijden moving activity no onvoltooid none none none informal adverbial ⟩
7204	⟨abstract subject veranderen cultural gradualCompletion no onvoltooid none precise none formal complement ⟩
6496	⟨speaker agent fietsen moving activity no onvoltooid none none none informal adverbial ⟩
6000	⟨abstract subject verminderen evaluation gradualCompletion no onvoltooid none none none formal complement ⟩
5977	⟨abstract subject veranderen cultural gradualCompletion no onvoltooid abstract none none formal complement ⟩
5692	⟨abstract subject ontstaan cultural gradualCompletion no onvoltooid none vague none formal main ⟩
5648	⟨hearer agent koken stativeMotion activity no onvoltooid none none none informal main ⟩
5604	⟨abstract subject gebeuren generic activity no onvoltooid none none none formal complement ⟩
5600	⟨abstract subject groeien stativePhysical gradualCompletion no onvoltooid none none none formal relative ⟩
5568	⟨hearer agent koken stativeMotion activity no onvoltooid none none none informal adverbial ⟩
5508	⟨human subject worden body gradualCompletion no onvoltooid none none none informal main ⟩

Table 6.24: The twenty most frequently used exemplars of the ah-progressive.

Chapter 7

Synthesis and final remarks

7.1 Construction grammar and Analogical Modeling

In this thesis, I have shown the similarity between the formal model of Analogical Modeling (Skousen 1989) on the one hand and usage-based construction grammar (Langacker 1987, Tomasello 2003a, Goldberg 2006) on the other. Whereas we can derive many good ideas from other formal, data-oriented models of categorization, many of them have presuppositions that conflict with those of usage-based construction grammar. Analogical Modeling is essentially an exemplar-based model, but the supracontextual schemas, derived from these exemplars in order to categorize new items, strongly resemble nodes in a constructional network. Because AM can model only one part of linguistic behavior (forced-choice categorization with a limited set of solutions), it is not equal to a fully operational model of construction grammar. To attain this, mechanisms of parsing and recombination would be required. Interesting advances in this direction can be found in Bannard, Lieven & Tomasello's (2009) Bayesian approach to learning function-based PCFGs as well as Bod's (2009) Unsupervised Data-Oriented Parsing and Beekhuizen's (2010) functional adaptation of this model.

Furthermore, a formal perspective on the cognitive entities assumed by construction grammar helped us criticize these assumptions. The tendency in construction grammar, and cognitive and functional approaches in general, is to attribute any interesting aspect of our mental life to some low level cognitive entity rather than to consider the possibility of this aspect being an emergent epiphenomenon. My claim is that we should be careful not to 'overload' the memory representation (which the storage of exemplars and schemas essentially is) with polysemy and metaphor links as well as statements which exemplars and schemas are prototypical. I believe entities such as these, however theoretically significant they are for cognitive approaches, should either be located at a higher, more holistic level of cognitive processes and/or be attributed to the analytical, explanatory perspective of the linguist investigating the construction. To my mind, all that happens in categorization is that cognitive agents use exemplars and schemas extracted from these to categorize new items. Other entities than these have no place in a model of categorization. This bold claim should be read as an invitation to those who do believe other entities have a place in categorization processes to show so in an off-line or on-line experimental setting.

Another consequence of superimposing the conceptual framework of AM on grammatical constructions is that the model truly becomes a full-entry model. Whereas in Construction Grammar constructions are stored parsimonously, and in Goldbergian construction grammar according to their entrenchment, the version of construction grammar advocated here states that all schemas are in principle extracted, but are not always equally useful because they are not always coherent, good predictors of a category label. The usage-based frequency aspect resides in the fact that these supracontextual schemas are weighted according to the cluster strength (the agreement measure) of the exemplars that underly them. Strikingly, this radical full-entry take on the constructicon is not something unheard of in the constructionist approaches. Langacker (2009) makes similar suggestions about the extraction of schemas in processing exemplars, which he claims to be always immanent, but not always put to use. The effect of the organization of AM is approximately the same, and thus can be said to be well-grounded in cognitive linguistic theorizing.

7.2 The question of abstraction

The parallelism between constructional and supracontextual schemas formed the starting point of the research executed in this thesis. The formal model thus can be used beyond its mere categorization capacities: we can show which schemas and exemplars are typical of a category (qualitative analysis of chapters 4 and 6) as well as model psycholinguistic findings (chapter 5). Analyses such as these show that formal modeling can also enhance our insight in the description or processing of grammatical structures.

The central issue investigated with AM is that of the degree of abstraction in constructional knowledge. If speakers extract schemas from the exemplars they process, as construction grammarians like to believe, up to what level of abstraction do they do so? And do they keep track of all concrete templates of grammatical knowledge? These two subquestions were operationalized by parametrizing one aspect of the model of AM, namely the abstraction of the schemas used. The abstraction of a schema was defined by the number of unspecified features it displayed. A schema that only has one open place thus has an abstraction $\alpha = 1$. With this information, we can exclude supracontextual schemas that fall within certain ranges of abstraction (R_α). Leaving out, for instance, ranges of supracontexts with low α-values amounts to asking the question whether the model can categorize adequately without detailed, specific schemas, only relying on abstract, more global rules. Leaving out supracontexts with high α-values amounts to asking the question whether the model can categorize without those global rules. With α being an integer between 0 and n, with n being the number of features coded, we can easily calculate all possible ranges of abstraction for a data set and assess the predictive power of the model at each one of them.

This predictive power was tested by letting the model predict the observed behavior of speakers in a corpus. The accuracy of the model thus is measured on the basis of actual language usage: the more instances of the speaker's categorizations the model predicts correctly, the more likely it is to be an appropriate mental model. This accuracy was measured by submitting a data set to a leave-one-out cross-validation test, and predicting for the test item at hand the outcome for each different settings of R_α. For many test items, it does not matter whether we use abstract schemas or concrete templates, but for some it does. More generally, there seem to be whole categories that are greatly effected by the exclusion of certain ranges of supracontexts.

Two sets of ranges are of special significance to the analysis, viz. that of $R_\alpha = [0 \ldots 0]$ to $R_\alpha = [0 \ldots n]$ and that of $R_\alpha = [0 \ldots n]$ to $R_\alpha = [n \ldots n]$. In the former set, we keep the minimum level of abstraction constantly at $\alpha = 0$, but we increase the maximum level of abstraction stepwise, thereby allowing slightly more abstract schemas with each run. What we can investigate with this set, is what the effect is of allowing more abstraction to the knowledge of our agent. The second set investigates the inverse: by stepwise increasing the minimum level of abstraction, we exclude more and more concrete schemas. Doing so, we can investigate how well a model predicts the language behavior without concrete knowledge.

7.2.1 Increasing abstraction

As to the former question (*What is the effect of adding abstraction to the representation?*), the results for the nine categories investigated are summarized in table 7.1. As we can see, most peaks lie halfway on the scale. Most models perform best if we allow maximally half the number of features to be open in schemas. Furthermore, for most categories, the performance approaches its asymptote much earlier: for the object and subject relative clauses at $R_\alpha = [0 \ldots 4]$ ($F = 0.974$) and $R_\alpha = [0 \ldots 5]$ ($F = 0.967$) respectively, and for the `sta` and `zit`-progressives at $R_\alpha = [0 \ldots 2]$ ($F = 0.186$) and $R_\alpha = [0 \ldots 3]$ ($F = 0.660$) respectively. One thing is sure: very abstract rules do not improve the performance of the model. Supracontexts with only, say, two or three values specified, do not turn false predictions into correct ones. This may be because many of the supracontexts at this level are heterogeneous or because the more concrete supracontexts 'inheriting' from this hypothetical abstract pattern already point in the right direction, and an abstract reinforcement is not needed. A more radical interpretation would be to say that if the curves approach their asymptote so early, as most do, speakers do not abstract beyond the very concrete (say, two or three open slots per pattern), and keep a memory base of large clouds of such schemas, that have an emergent effect of stability and generalizability.

Three cases form interesting deviations from this 'steep rise to the asymptote' pattern found in six of the nine categories, viz the `PP`-dative, the `lig`-progressive and the `loop`-progressive. In the former two, we find that there is a clear peak, and that after the peak the score drops somewhat to an asymptote, where it will remain un-

category	peak		$R_\alpha = [0 \dots n]$	shape curve
	R_α	F-score	F-score	
`NP`	$[0 \dots 5]$	0.903	0.903	steep, asymptotic
`PP`	$[0 \dots 3]$	0.422	0.370	steep, drops after peak
`object`	$[0 \dots 8]$	0.977	0.977	steep, asymptotic
`subject`	$[0 \dots 8]$	0.974	0.974	steep, asymptotic
`ah`	$[0 \dots 6]$	0.669	0.664	steep, asymptotic
`lig`	$[0 \dots 6]$	0.384	0.366	varied
`loop`	$[0 \dots 8]$	0.210	0.206	gradual rise, asymptotic
`sta`	$[0 \dots 9]$	0.190	0.190	steep, asymptotic
`zit`	$[0 \dots 6]$	0.668	0.668	steep asymptotic

Table 7.1: The effect of increasing abstraction over the nine categories investigated.

till all abstraction is allowed. This pattern is most saliently present in the `PP`-dative, with a difference of 0.052 between the peak and the asymptote, and less clearly so in the `lig`-progressive, where the difference in F-scores is only 0.018. Both cases are minority categories, that only constitute a small percentage of the total. It seems that the only way that they can maintain stability, is if there are no abstract schemas applicable to them, derived from other, majority, patterns (such as the `NP`-dative and the `zit`-progressive). Such abstract schemas will worsen the model's performance by overruling the constellation of concrete schemas making the correct prediction.

On the other hand, the `loop`-progressive seems to be a case in which more abstract schemas continue to help the model's performance to improve. The gradual increase means that, as more abstract schemas are allowed, the pattern is more often correctly predicted. This might point to the fact that the `loop` progressive is not a very broad construction, that shows much diversity in its meaning, as the other progressives do, but has a limited, narrow band of meanings, such as 'communicating in an annoying way', 'doing something repetitively (in an annoying way)'. The qualitative analysis in chapter 6 shows that the lexical meaning of 'walk' does not surface in a related grammatical meaning of 'doing something while moving' or something like that. We must keep in mind, however, that with the current variables, we can predict only about 20% of the variation, so that all of the previous statements (if not the entire thesis) should be read with the caveat 'given the current data set and variable set'.

7.2.2 Decreasing specificity

Then there is the question what the effect of decreasing the allowed concreteness of supracontextual schemas is. What we see, is that for majority categories that can be very well predicted, the allowed minimum level of abstraction is very high. An overview of the results can be found in table 7.2. For the `NP`-dative, the `object` and the `subject` relative clauses, the prediction deteriorates only if we allow just the most abstract schemas. For some other categories, there is a gradual decrease, with the model predicting a few percent worse each time the next level of specificity or concreteness is disallowed. This points to the fact that the ability to generalize, especially for smaller patterns, such as the `lig`-progressive and the `PP`-dative, lies not only in having sufficiently open-ended patterns, but also in having sufficiently concrete patterns that can dominate certain patches of the conceptual space in which the constructions are located. For all cases, there is a central group of items that will be predicted correctly, given any non-extreme range of allowed supracontexts, but the 'periphery' of the category is lost given a model that does not employ concrete and detailed linguistic knowledge.

An interesting observation is the fact that none of the categories performs best if all very concrete supracontexts are allowed, as the peaks are always at ranges with a lower bound of $\alpha = 3$. In none of the cases, however, the peak is much higher than the score at $R_\alpha = [0 \dots n]$, where n is the number of variables. The biggest differences in F-scores between the model with all specificity allowed and the peak performance on the decreasing specificity scale are for the `loop`-progressive and the `sta`-progressive, namely some 0.02. For most other cases, the difference is smaller than one percent.

These two patterns thus seem to behave somewhat differently. For the `loop`-progressive and the `sta`-progressive, there is a true rise in performance if concrete schemas are excluded from the categorization task. Does this mean that concrete knowledge harms the performance of the model for some categories? I believe that for the `sta`-progressive this is hard to say, as the performance on the category in general lies just above the baseline level, which means that the model does not perform much better than blind chance given the variables and dataset we are currently using. For the `loop`-pattern, I believe there might be some truth to the fact that this construction needs abstraction to perform well, as I explained in section 7.2.1, but also to the fact that very concrete schemas seem to 'take away' correct predictions from this pattern. The coherence of this category makes it more strongly represented among the abstract schemas, because the abstract supracontextual schemas are often homogeneous and supported by many exemplars, and at the same time less strongly represented among the concrete ones, as they only cover a relatively narrow range of values because of the same coherence.

	peak		first drop	
category	R_α	F-score	R_α	shape
NP	$[3 \dots 9]$	0.904	$[7 \dots 9]$	long sustain, deep drop
PP	$[3 \dots 9]$	0.374	$[4 \dots 9]$	gradual decline
object	$[7 \dots 11]$	0.979	$[10 \dots 11]$	long sustain, deep drop
subject	$[7 \dots 11]$	0.976	$[9 \dots 11]$	long sustain, deep drop
ah	$[3 \dots 12]$	0.667	$[7 \dots 12]$	very gradual decline
lig	$[7 \dots 12]$	0.373	$[8 \dots 12]$	gradual decline
loop	$[7 \dots 12]$	0.231	$[8 \dots 12]$	gradual rise, gradual decline
sta	$[7 \dots 12]$	0.211	$[11 \dots 12]$	gradual rise, deep drop
zit	$[3 \dots 12]$	0.670	$[6 \dots 12]$	gradual decline

Table 7.2: The effect of decreasing specificity over the nine categories investigated.

7.3 Some remaining issues and a look ahead

Methodological studies like the present one can help develop the cognitivist and constructionist frameworks. Further steps with methodologies like the one presented here can be taken by bringing together expertise from domains such as cognitive psychology, mathematical psychology, machine learning, artificial intelligence, natural language processing and speech technology, psycholinguistics and, obviously, linguistics itself. In this thesis, I only touched upon a small subset of what we already know, but most knowledge still is floating around in papers I have not read yet and lingering in subdisciplines I have not explored yet. However, only with a small paralellism such as the one between supracontextual schemas in AM and constructions in usage-based construction grammar, and inspired by insights from mathematical psychology (the Varying Abstraction Model), we can get ahead pretty quickly. The results do not only sharpen our understanding of how language works cognitively, but may also provide us with valuable insights in the nature of linguistic category learning.

In the light of the findings in this model, there is the question how speakers know what the optimal range of abstraction is. I have not dealt with this issue before, but I think this amounts to a parameter setting problem that is dealt with on an individual basis and within individuals, per category, or per set of mutually exclusive categories, as is the case in alternations. Somehow, the language user must decide which ranges of supracontexts to use in classifying new items, and I firmly believe that this process is data-driven as well, especially given the diversity we have seen in the optimum ranges of abstraction per category.

However, a parameter setting does not mean that whatever is outside of the selected range is forgotten immediately. The schemas are in the end not true mental objects, but perhaps weights on links between memory points. As such, they may exist only in the organization of the exemplar network. The network of schemas then becomes a metaphor for a neural organizational structure that is even harder to grasp for us. As Langacker (2009) states: the schemas are immanent in the exemplars, and I interpret this to mean that they can always 'be found' in the exemplars, but that they are not autonomous entities in the memory. So if the language user remembers the exemplars, that consist of a string of features, all schemas can in principle always be 'highlighted' temporarily and used for classification. The parameter setting of the range of abstraction most likely used determines in that case what schemas are more likely to be central to the categorization process.

Another issue is that of the empirical coverage of the determinants. If the set of determinants only reach F-scores of around $F = 0.40$, is the development over the different levels of abstraction still valid? It might be that the abstraction curves change drastically if we added one variable that singlehandedly raises the F-score to $F = 0.60$, which is not impossible. That uncertainty being as it is, we do find recurrent patterns in the performance of the model at different ranges. Especially salient is the fact that an increase of abstraction typically has little effect beyond $\alpha = 4$, irrespective of whether the right set of determinants has been found.

A final, practical, issue is that of the selection of features. In phonological analogy, the variables are the phonemes, aligned to each other in one way or the other, or possibly phonological features, such as place of articulation. For functional analogy, which formed the basic mechanism behind the studies in this thesis, the set of features is somewhat more arbitrary. This is not problematic, as the model should ideally be able to find the useful, homogeneous supracontexts and more or less 'ignore' the noise, so that we can add more variables freely. It is expected that as more variables are added, the model will not perform worse. Whether this is in practice the case, has not been investigated here. The absolute exclusion of variables touches on important questions concerning the selective attention humans have to certain variables for certain phenomena and how this bias emerges. The theory of AM does not yet provide a satisfying answer to this.

Some more abstract questions remain unanswered as well, and this is the best place to discuss them one last time. Firstly, feature-based exemplar models, such as AM and Memory-Based Learning, already start from parsed representations of exemplars. The exemplars of chairs and stool are thus each composed out of an arms-feature, a back-feature and a legs-feature. But how do these parsings emerge in the first place out of the entities that are perceived holistically at first (unless, of course, we assume categories to be innate). And how does a cognitive agent, given that she can parse an object into features, select a finite set of features out of the potentially infinite number of aspects she might attend to?

These questions might seem rather 'philosophical', but they are also the kind of criticisms found in critiques of functional approaches. How do you limit the hypothesis space so that the computation does not become unrealistic? How do you know what are the right constituents? Answers to these questions are not provided in this study, but I believe we should look in the direction of the social setting in which learning and using language is always embedded (cf. Clark 1996). As Tomasello (2003a) argues, the joint attentional frame shared by caregiver and child is paramount for bootstrapping the hypothesis space of the child. Only through interacting with other language users and the objects of the world, can the child zero in on what is salient for members of her community to communicate about. What the models of associative learning behind these learning techniques are, on the level of 'lonely cognition', is an issue to which I hope to have contributed with this study.

Bibliography

Alishahi, A. & S. Stevenson (2008). 'A computational model of early argument structure acquisition'. *Cognitive Science*, **32**(5): 789–834.

Ariel, M. (2001). 'Accessibility theory: an overview'. In: T. Sanders, J. Schilperoord & W. Spooren, eds., *Text Representation*. Philadelphia/Amsterdam: John Benjamins, 29–87.

Bannard, C., E. Lieven & M. Tomasello (2009). 'Modeling children's early grammatical knowledge'. *Proceedings of the National Academy of Sciences*, **106**(41): 17,284–17,289.

Barsalou, L. W. (1992). *Cognitive psychology: an overview for cognitive scientists*. Hillsdale, NJ: Lawrence Erlbaum.

Beach, L. R. (1974). 'Cue probabilism and inference behavior'. *Psychological Monographs: General and Applied*, **78**: 1–20.

Van der Beek, L. (2004). 'Argument order alternations in Dutch'. In: M. Butt & T. H. King, eds., *Proceedings of the LFG04 Conference. University of Canterbury*. Stanford, CA: CSLI, 39–59.

Beekhuizen, B. (2010). 'Constructionist DOP: A computational take on construction learning'. Tutorial paper, University of Amsterdam.

Bod, R. (1998). *Beyond grammar: An experience-based theory of language*. Stanford, CA: CSLI.

Bod, R. (2006). 'Unsupervised parsing with U-DOP'. In: L. Màrquez & D. Klein, eds., *Proceedings of the Tenth Conference on Computational Natural Language Learning*. 85–92.

Bod, R. (2009). 'Constructions at work or at rest?' *Cognitive Linguistics*, **20**(1): 129–134.

Boogaart, R. J. (1999). *Aspect and temporal ordening. A contrastive analysis of Dutch and English*. Ph.D. thesis, Vrije Universiteit Amsterdam.

Booij, G. E. (2008). 'Constructional idioms as products of linguistic change: the *aan het* + INFINITIVE construction in dutch'. In: A. Bergs & G. Diewald, eds., *Constructions and language change*. Berlin: Mouton de Gruyter, 79–104.

Borsley, R. D. & F. J. Newmeyer (2009). 'On subject-auxiliary inversion and the notion purely formal generalization'. *Cognitive Linguistics*, **20**(1): 135–143.

Bresnan, J., A. Cueni, T. Nikitina & H. Baayen (2007). 'Predicting the dative alternation'. In: G. Boume, I. Kraemer & J. Zwarts, eds., *Cognitive Foundations of Interpretation*. Amsterdam: Royal Netherlands Academy of Science, 69–94.

Bybee, J. (2006). 'From usage to grammar: The mind's response to repetition'. *Language*, **82**(2): 711–733.

Chandler, D. & R. Skousen (1997). 'Analogical modeling and the English past tense: A reply to Jaeger et al. 1996'. Paper retrieved from http://humanities.byu.edu/am/am-biblio.html.

Chapman, D. & R. Skousen (2005). 'Analogical modeling and morphological change: The case of the adjectival negative prefix in English'. *English Language and Linguistics*, **9**(2): 1–25.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Colleman, T. (2006). *De Nederlandse datiefalternantie. Een constructioneel en corpusgebaseerd onderzoek.* Ph.D. thesis, Ghent University.

Crain, S., R. Thornton & D. Khlentzos (2009). 'The case of the missing generalizations'. *Cognitive Linguistics*, **20**(1): 145–155.

Daelemans, W. & A. van den Bosch (2005). *Memory-Based Language Processing*. Cambridge: Cambridge University Press.

Delorge, M. (2009). *De relatie tussen betekenis en structuur bij privatieve en receptieve werkwoorden in het Nederlands*. Ph.D. thesis, Universiteit Gent.

Derwing, B. & R. Skousen (1994). 'Productivity and the English past tense: Testing Skousen's analogy model'. In: S. D. Lima, R. L. Corrigan & G. K. Iverson, eds., *The Reality of Linguistic Rules*. Amsterdam/Philadelphia: John Benjamins, 193–218.

Givón, T. (1983). 'Topic continuity in discourse: an introduction'. In: T. Givón, ed., *Topic continuity in discourse: a quantitative cross-language study*. vol. 3 of *Typological Studies in Language*. Amsterdam/Philadelphia: John Benjamins, 1–42.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.

Gries, S. T. (2003). *Multifactorial analysis in corpus linguistics: a study of Particle Placement*. London, New York: Continuum Press.

Gries, S. T. & A. Stefanowitsch (2004). 'Extending collostructional analysis: a corpus-based perspective on "alternations"'. *International Journal of Corpus Linguistics*, **9**(1): 97–129.

Grondelaers, S. (2009). 'Woordvolgorde in presentatieve zinnen en de theoretische basis van multi-factoriële grammatica'. *Nederlandse Taalkunde*, **14**(3): 282–299.

Grondelaers, S., K. Deygers, H. van Aken, V. van den Heede & D. Speelman (2000). 'Het CONDIV-corpus geschreven nederlands'. *Nederlandse Taalkunde*, **5**(4): 356–363.

Haeseryn, W., K. Romijn, G. Geerts, J. de Rooij & M. van den Toorn (1997). *Algemene Nederlandse Spraakkunst*. Groningen/Deurne:Martinus Nijhoff, 2nd edn.

Hopper, P. & S. A. Thompson (1980). 'Transitivity in grammar and discourse'. *Language*, **56**: 251–299.

Israel, M., C. R. Johnson & P. Brooks (2000). 'From states to events: The acquisition of English passive participles'. *Cognitive Linguistics*, **11**(1/2): 1–27.

Kay, P. & C. J. Fillmore (1999). 'Grammatical constructions and linguistic generalizations: the *What's X doing Y?* construction'. *Language*, **75**: 1–33.

Lakoff, G. (1987). *Women, Fire and Dangerous Things. What categories reveal about the mind*. Chicago: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar.*, vol. I: Theoretical Prerequisites. Stanford: Stanford University Press.

Langacker, R. W. (2009). 'A dynamic view of usage and language acquisition'. *Cognitive Linguistics*, **20**(3): 627–640.

Langacker, R. W. (1991). *Foundations of Cognitive Grammar*, vol. II: Descriptive Application. Stanford: Stanford University Press.

Lee, M. D. & W. Vanpaemel (2008). 'Exemplars, prototypes, similarities and rules in category representation: An example of hierarchical Bayesian analysis'. *Cognitive Science*, **32**: 1403–1424.

Lemmens, M. (2005). 'Aspectual posture verb constructions in Dutch'. *Journal of Germanic Linguistics*, **17**(3): 183–217.

Mak, W. M., W. Vonk & H. Schriefers (2002). 'The influence of animacy on relative clause processing'. *Journal of Memory and Language*, **47**: 50–68.

Mak, W. M., W. Vonk & H. Schriefers (2006). 'Animacy in processing relative clauses: The hikers that rocks crush'. *Journal of Memory and Language*, **54**: 466–490.

Medin, D. L. & M. M. Schaffer (1978). 'Context theory of classification learning'. *Psychological Review*, **85**(3): 207–238.

Ney, H., S. Martin & F. Wessel (1997). 'Statistical language modeling using leaving-one-out'. In: S. Young & G. Bloothooft, eds., *Corpus-based Methods in Language and Speech Processing*. Dordrecht: Kluwer Academic, 174–207.

Posner, M. I. & S. W. Keele (1968). 'On the genesis of abstract ideas'. *Journal of Experimental Psychology*, **77**(3/1): 353–363.

Rosch, E. (1978). 'Principles of categorization'. In: E. Rosch & B. B. Lloyd, eds., *Cognition and Categorization*. chap. 2. Hillsdale, NJ: Lawrence Erlbaum, 27–48.

Rosch, E. & C. B. Mervis (1975). 'Family resemblances: Studies in the internal structure of categories'. *Cognitive Psychology*, **7**(4): 573–605.

Ross, B. H. & V. S. Makin (1999). 'Prototype versus exemplar models in cognition'. In: *The Nature of Cognition*. chap. 7. Cambridge, MA: MIT Press, 205–241.

Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt, Brace.

Silverstein, M. (1976). 'Hierarchy of features and ergativity'. In: R. Dixon, ed., *Grammatical categories in Australian languages*. Canberra: Humanities Press, 112–171.

Skousen, R. (1989). *Analogical Modeling of Language*. Dordrecht: Kluwer.

Skousen, R. (2002). 'An overview of Analogical Modeling'. In: *Analogical Modeling. An exemplar-based approach to language*. Amsterdam/Philadelphia: John Benjamins, 11–26.

Skousen, R., D. Lonsdale & D. B. Parkinson, eds. (2002). *Analogical Modeling: An exemplar-based approach to language*, vol. 10 of *Human Cognitive Processing*. Amsterdam/Philadelphia: John Benjamins.

Taylor, J. R. (1998). 'Syntactic constructions as prototype categories'. In: M. Tomasello, ed., *The new psychology of language. Cognitive and functional approaches to language structure*. vol. 1, chap. 7. Mahwah: Erlbaum, 177–202.

Tomasello, M. (2003a). *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M., ed. (2003b). *The new psychology of language. Cognitive and functional approaches to language structure*, vol. 2. Mahwah: Erlbaum.

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Vendler, Z. (1957). 'Verbs and times'. *The Philosophical Review*, **66**(2): 143–160.

Verbeemen, T., W. Vanpaemel, S. Pattyn, G. Storms & T. Verguts (2007). 'Beyond exemplars and prototypes as memory representations of natural concepts: A clustering approach'. *Journal of Memory and Language*, **56**: 537–554.

Verhagen, A. (2009). 'The conception of constructions as complex signs. Emergence of structure and reduction to usage'. *Constructions and Frames*, **1**(1): 119–152.

Vismans, R. (1982). '"Durative" constructions in modern Dutch'. In: A. Fry, J. Mackenzie & R. Todd, eds., *Free University Studies in English*. vol. 2. Amsterdam: Free University, 243–268.

Wittgenstein, L. (1963 (1953)). *Philosophische Untersuchungen*. Frankfurt a/M: Suhrkamp.

Zaenen, A., J. Carletta, G. Garretson, J. Bresnan, A. Koontz-Garboden, T. Nikitina, M. C. O'Connor & T. Wasow (2005). 'Animacy encoding in English: Why and how'. In: *Proceedings of the 2004 ACL Workshop on Discourse Annotation*. 118–125.

Zubin, D. A. (1979). 'Discourse function of morphology: The focus system in German'. In: T.Givón, ed., *Syntax and semantics:*. vol. 12 of *Discourse and syntax*. New York: Academic Press, 469–504.

Appendix A

CD with data

This CD contains all data sets used in this thesis as well as the thesis itself in pdf-format. The material found on the CD can also be found on the website accompanying this thesis:
sites.google.com/site/barendbeekhuizen/thesis