

Computational implementation of a linear phase parser. Framework and technical
documentation
(version 19)

2019

(Revised April 2024)

Pauli Brattico

Abstract

This document describes a computational implementation of a linear phase grammar. The model assumes that the core computational operations of narrow syntax are applied incrementally on a phase-by-phase basis in language comprehension. Full formalization with a description of the Python implementation will be discussed, together with a few words towards empirical justification. The theory is assumed to be part of the human language faculty (UG).

How to cite this document: Brattico, P. (2019). Computational implementation of a linear phase parser. Framework and technical documentation (version 19.0). IUSS, Pavia.

1	INTRODUCTION	7
2	INSTALLATION AND USE	8
2.1	INSTALLATION	8
2.2	USE.....	9
2.3	REPLICATION.....	11
3	FRAMEWORK	12
3.1	THE FRAMEWORK.....	12
3.2	COMPUTATIONAL GENERATIVE GRAMMAR FRAMEWORK	17
4	THE CORE SYNTACTIC ENGINE	21
4.1	INTRODUCTION.....	21
4.2	MERGE-1.....	21
4.3	LEXICAL SELECTION FEATURES	26
4.4	PHASES AND LEFT BRANCHES	28
4.5	LABELING	30
4.6	UPWARD PATHS AND MEMORY SCANNING	32
4.7	ADJUNCT ATTACHMENT.....	34
4.8	TRANSFER.....	37
4.8.1	<i>Introduction.....</i>	37
4.8.2	<i>Phrasal reconstruction: general comments.....</i>	38
4.8.3	<i>Ā-chains.....</i>	40
4.8.4	<i>A-chains.....</i>	43
4.8.5	<i>Head reconstruction</i>	45
4.8.6	<i>Adjunct reconstruction (also scrambling reconstruction).....</i>	48
4.8.7	<i>Minimal search</i>	49
4.8.8	<i>AgreeLF.....</i>	50
4.8.9	<i>Ordering of operations</i>	52

4.9	LEXICON AND MORPHOLOGY	53
4.9.1	<i>From phonology to syntax</i>	53
4.9.2	<i>Lexical items and lexical redundancy rules</i>	56
4.9.3	<i>Derivational and inflectional morphemes</i>	58
4.9.4	<i>Lexical features</i>	58
4.10	NARROW SEMANTICS.....	59
4.10.1	<i>Syntax, semantics and cognition</i>	59
4.10.2	<i>Argument structure</i>	63
4.10.3	<i>Predicates, arguments and antecedents</i>	65
4.10.4	<i>The pragmatic pathway</i>	66
4.10.5	<i>Operators</i>	67
4.10.6	<i>Binding</i>	68
5	PERFORMANCE	70
5.1	HUMAN AND ENGINEERING PARSERS.....	70
5.2	MAPPING BETWEEN THE ALGORITHM AND BRAIN	70
5.3	COGNITIVE PARSING PRINCIPLES.....	71
5.3.1	<i>Incrementality</i>	71
5.3.2	<i>Connectness</i>	75
5.3.3	<i>Seriality</i>	75
5.3.4	<i>Locality preference</i>	76
5.3.5	<i>Lexical anticipation</i>	76
5.3.6	<i>Left branch filter</i>	77
5.3.7	<i>Working memory</i>	77
5.3.8	<i>Conflict resolution and weighting</i>	77
5.4	MEASURING PREDICTED COGNITIVE COST OF PROCESSING	78
5.5	A NOTE ON IMPLEMENTATION	79
6	INPUTS AND OUTPUTS	80
6.1	INSTALLATION AND USE	80

6.2	STRUCTURE OF THE INPUT FILES	81
6.2.1	<i>Study configuration file</i>	81
6.2.1.1	General.....	81
6.2.1.2	Metadata.....	82
6.2.1.3	Files	82
6.2.1.4	General simulation parameters.....	83
6.2.1.5	Image parameters	83
6.2.1.6	UG components	84
6.2.1.7	Parsing heuristics	84
6.2.2	<i>Test corpus file (any name)</i>	84
6.2.3	<i>Lexical files</i>	87
6.2.4	<i>Lexical redundancy rules</i>	88
6.3	STRUCTURE OF THE OUTPUT FILES.....	89
6.3.1	<i>Errors.....</i>	89
6.3.2	<i>Results</i>	90
6.3.3	<i>The derivational log file.....</i>	90
6.3.4	<i>Simple logging.....</i>	93
6.3.5	<i>Resources file.....</i>	93
6.3.6	<i>Phrase structure images.....</i>	93
6.4	INPUTS, OUTPUTS AND THE GRAPHICAL USER INTERFACE	94
7	GRAMMAR FORMALIZATION.....	95
7.1	BASIC GRAMMATICAL NOTIONS (PHRASE_STRUCTURE.PY)	95
7.1.1	<i>Introduction.....</i>	95
7.1.2	<i>Lexical items</i>	95
7.1.3	<i>Phrase structure geometry</i>	95
7.1.4	<i>Sisters</i>	96
7.1.5	<i>Proper selected complement</i>	96
7.1.6	<i>Heads (labels), maximal projection, container.....</i>	97
7.1.7	<i>Minimal search, geometrical minimal search and upstream search.....</i>	98

7.1.8	<i>Upward paths</i>	99
7.1.9	<i>Cyclic Merge</i>	100
7.1.10	<i>Countercyclic Merge-1</i>	100
7.1.11	<i>Remove</i>	101
7.1.12	<i>Feature checking</i>	102
7.1.13	<i>Probe-goal: probe(label, goal_feature)</i>	102
7.1.14	<i>Tail-head relations</i>	103
7.2	TRANSFER.....	104
7.2.1	<i>A comment on the current version (15.0)</i>	104
7.2.2	<i>Generalized transfer</i>	104
7.2.3	<i>Chain creation (all chains)</i>	106
7.2.4	<i>Head chain</i>	108
7.2.5	<i>A-chains</i>	109
7.2.6	<i>Ā-chains</i>	109
7.2.7	<i>Agree</i>	110
7.2.8	<i>Scrambling (scrambling_reconstruction.py)</i>	111
7.2.9	<i>Adjunct promotion (adjunct_constructor.py)</i>	111
7.3	LF-LEGIBILITY (LF.PY)	112
7.3.1	<i>Overall</i>	112
7.3.2	<i>Selection tests</i>	114
7.3.3	<i>Projection principle</i>	114
7.4	SEMANTICS (NARROW_SEMANTICS.PY)	115
7.4.1	<i>Introduction</i>	115
7.4.2	<i>Projecting semantic inventories (semantic switchboard)</i>	115
7.4.3	<i>Quantifiers-numerals-denotations module</i>	116
7.4.4	<i>Operator-variable interpretation (SEM_operators_variables.py)</i>	117
7.4.5	<i>Pragmatic pathway</i>	118
7.4.6	<i>Argument-predicate pairs</i>	120

8 FORMALIZATION OF THE PARSER	122
8.1 LINEAR PHASE PARSER (LINEAR_PHASE_PARSER.PY)	122
8.1.1 <i>The speaker model</i>	122
8.1.2 <i>Parse sentence (parse_sentence)</i>	123
8.1.3 <i>Recursive parsing function (parse_new_item)</i>	123
8.2 PSYCHOLINGUISTIC PLAUSIBILITY	128
8.2.1 <i>General architecture</i>	128
8.2.2 <i>Filtering</i>	129
8.2.3 <i>Ranking (rank_merge_right_)</i>	130
8.2.4 <i>Knocking out heuristic principles</i>	132
8.3 RESOURCE CONSUMPTION	132
9 GRAPHICAL USER INTERFACE.....	135
9.1 LAUNCHING THE INTERFACE AND LOADING STUDIES.....	135
9.2 ANALYZING INDIVIDUAL SENTENCES	136
9.3 RUNNING ONE COMPLETE STUDY.....	139
9.4 EXAMINING THE ROOT LEXICONS AND SPEAKER MODELS.....	139
9.5 SIMULATION PARAMETERS AND SETTINGS	140

1 Introduction

This document describes a computational Python-based implementation of a linear phase grammar that was originally developed and written by the author while working in an IUSS-funded research project between 2018-2020, in Pavia, Italy, and then continued as an independent project.¹ The algorithm is interpreted as a realistic description of the information processing steps involved in real-time language comprehension. It captures both cognitive principle of language (“competence”) and cognitive mechanisms involved in language comprehension and use (“performance”).

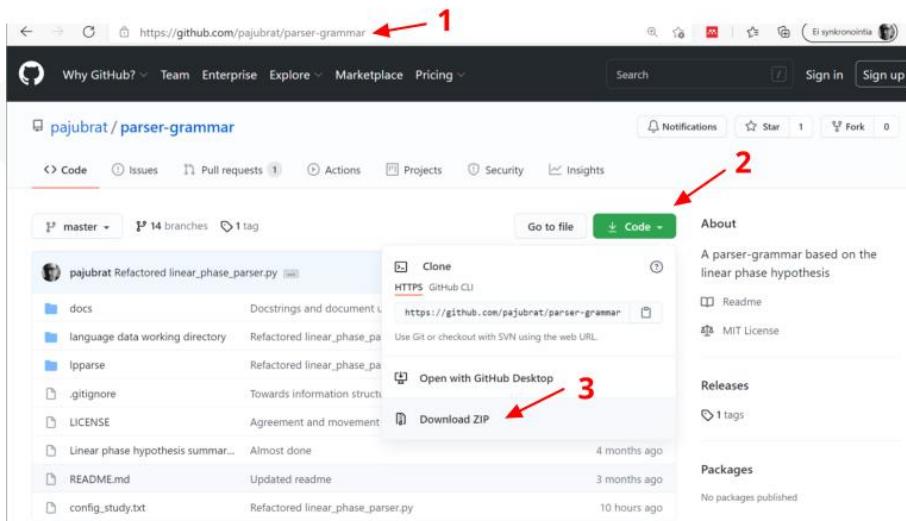
This document describes properties of the version 19, which keeps within the framework of the original work but provides improvements, corrections and additions. There may be mismatches between what is described here and what appears in the latest version of the source code. This is because the documentation lags behind and/or has not been completed due to the experimental nature of changes and additions that do not warrant documentation. In addition, some parts of this document are in better condition than others, and there are gaps awaiting documentation. This happens when the material is still being worked on and has not been accepted for publication.

¹ The research was conducted in part under the research project “ProGraM-PC: A Processing-friendly Grammatical Model for Parsing and Predicting Online Complexity” funded internally by IUSS (Pavia).

2 Installation and use

2.1 Installation

The linear phase algorithm is a collection of Python functions that processes natural language sentences by using principles of human linguistic competence and performance. It works by reading test sentences from a dataset (test corpus) file and analyzing them. The program can be installed on a local computer by cloning it from the source code repository <https://github.com/pajubrat/parser-grammar>. The easiest method is by downloading the software package as one ZIP file and extracting it into a directory in the local machine. Navigate to the source repository by using any web browser (1, Figure below), then click the button “Code” (2) and select “Download ZIP” (3).



Once the package is on the local machine, it must be unpacked into any directory. The same files and folders displayed on the above figure should appear in that directory. Because the script is written in Python (3x) programming language, the researcher must have Python installed on the

local machine.² After installation, the local installation directory should contain the following files and folders:

📁 .git	14.4.2024 12.34	Tiedostokansio
📁 .idea	15.4.2024 9.11	Tiedostokansio
📁 docs	10.6.2023 11.08	Tiedostokansio
📁 language data working directory	4.4.2024 10.39	Tiedostokansio
📁 lpparse	15.4.2024 9.39	Tiedostokansio
📝 \$app_settings.txt	15.4.2024 9.28	TXT-tiedosto 1 kt
📄 .gitignore	21.1.2024 11.24	GITIGNORE-tiedosto 1 kt
📝 dev_log.txt	15.4.2024 9.39	TXT-tiedosto 0 kt
📄 LICENSE	19.2.2022 10.33	Tiedosto 2 kt
📄 README.md	9.4.2024 20.42	MD-tiedosto 2 kt

Folder */docs* contains documentation (e.g., this document and previous versions), */language data working directory* stores the input and output files associated with any particular study, and */lpparse* stores the modules containing the program code. *\$app_settings.txt* is a text file that configures the script, in a manner explained below. In most cases the script comes preconfigured so that it runs some particular study and can therefore be tested without providing further information or parametrization.

2.2 Use

The script is launched by having the Python interpreter to run the program script. In most cases the version copied or cloned from the code repository comes in a working configuration, it should do something meaningful out of the box.³ To execute the script in Windows, start a command prompt program such as Windows PowerShell or the command prompt (“cmd”) and navigate to the local installation directory. The script can be executed by command *python lpparse*. If the

² Follow the instructions from www.python.org. In Windows, Python installation path must be provided in the Windows PATH environmental variable. Search for the term “setting windows PATH variables.”

³ The version downloaded by following the instructions above is always the latest. If the user wants to download older versions, say for replication purposes, then these can be downloaded by first selecting the branch and then performing the instructions above; see below.

local installation directory is **J:\Research** (as it is on the author's local machine), then the script is executed as follows:



```
Microsoft Windows [Version 10.0.19045.4291]
(c) Microsoft Corporation. Kaikki oikeudet pidätetään.

J:\Research>python lpparse
```

The program reads a study configuration file *\$app_settings.txt* from the local installation directory which points to the parameters used in the simulation trial. One of these parameters is the folder and name of the *test corpus file* that contains the sentences the script will analyze. For example:

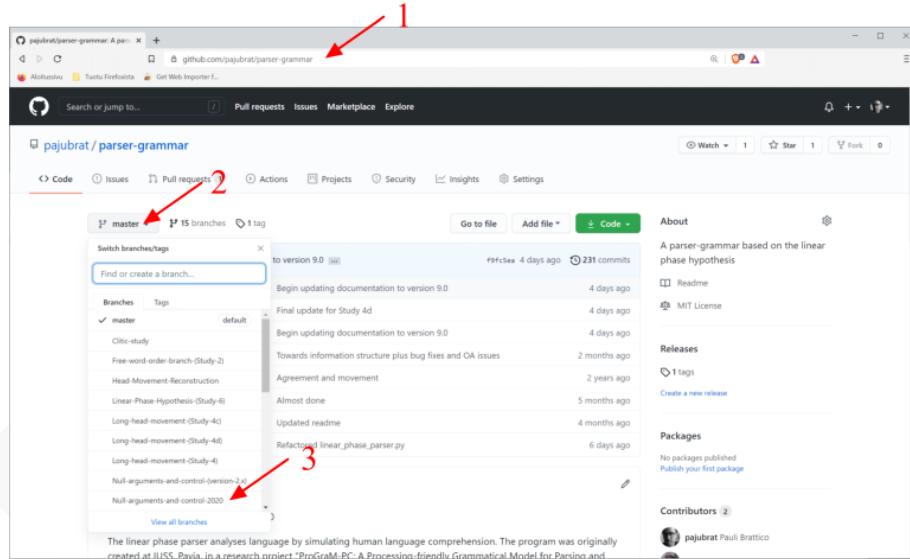
```
3  & 1. Finite agreement
4
5      & 1.1 Subject-finite element agreement
6
7          & 1.1.1 Subject-verb agreement
8
9              & First person singular
10
11                 minä lähdet
12                 'I.nom leave.1sg'
13
14                 minä tein työtä
15                 'I.nom make.pst.1sg work.par'
16
17              & Second person singular
18
19                 sinä lähdet
20                 'you.nom leave.2sg'
21
22                 sinä teit työtä
23                 'you.nom make.pst.2sg work.par'
```

This is a screenshot of the dataset (a list of sentences with comments) the script processes (the contents of the dataset file will obviously depend on the study). The results are generated into several *output files* inside the study folder. These output files record the results of the execution of the script and include such information as how many mistakes the script did, how did it derive various sentences, what type of semantic interpretations it associated with each sentence, and so on. An alternative way to use the program is to launch the graphical user interface by command *python lpparse -app*. This is described in Chapter 9. Some output is also generated to the command prompt when the script is calculating the dataset.

2.3 Replication

When a study is published, the input files plus its source code is stored in source code repository.

The source code repository (that runs on a program called Git) not only stores the current version of the script, but also maintains a record of previous versions. The development history can also be branched, meaning that it is possible to develop several versions of the software in parallel (and possibly merge them later). These parallel branches are currently used to store unambiguous snapshots of the scripts that were used in connection with published studies. To access them, navigate again to the source code repository (1, Figure below), then click for the tab shown in the figure below (2) and select the branch that is of interest (3).



This should select a snapshot of a development branch that contains a version of the script that was used in a published study. The name of the relevant branch is usually mentioned in the published paper. Use of these branches is not recommended for anything else than replication.

3 Framework

3.1 The framework

A native speaker can interpret sentences in his or her language by various means, for example, by reading sentences printed on a paper. Since it is possible accomplish this task without external information, all information required to interpret a sentence in one's native language must be present in the sensory input. We assume that efficient and successful language comprehension is possible from contextless and unannotated sensory input.⁴

Some sensory inputs map into meanings that are hard or impossible to construct (e.g., *democracy sleeps furiously*), while others are judged as ungrammatical. A useful theory of language comprehension must appreciate these properties. The model therefore defines a characteristic function from sensory objects into a binary (in some cases graded) classification of its input in terms of grammaticality or some related notion, such as semanticality, marginality or acceptability. These categorizations are studied by eliciting responses from native speakers. Any language comprehension model that captures this mapping correctly is said to be *observationally adequate*. Any scientific theory must be, minimally, observationally adequate, and the fact that it is observationally adequate must be shown by deductive calculation. Because the model judges input sentences, it implements a .

Some aspects of the comprehension model are language-specific, others are universal and depend on the biological structure of the human brain. A universal property can be elicited from speakers

⁴ Some aspects of semantic interpretation, such as the intended denotations of pronouns, depend on the context. If no context is present, the sentence receives an all-new reading.

of any language. For example, it is a universal property that an interrogative clause cannot be formed by moving an interrogative pronoun out of a relative clause (as in, e.g., **who did John met the person that Mary admires__?*). On the other hand, it is a property of Finnish and not English that singular marked numerals other than ‘one’ assign the partitive case to the noun they select (*kolme sukka-a* ‘three.sg.0 sock.sg.par’). The latter are acquired from the input during language acquisition. Universal properties plus the storage systems constitute the fixed portion of the grammar, whereas the language-specific contents stored into the memory systems during language acquisition and other relevant experiences constitute the language-specific, variable part. A theory of language comprehension that captures the fixed and variable components in a correct or at least realistic way without contradicting anything known about the process of language acquisition is said to reach *explanatory adequacy*.

It is possible to design an observationally adequate comprehension theory for Finnish such that it replicates the responses elicited from native speakers, yet the same model and its principles would not work in connection with a language such as English, not even when provided with a fragment of English lexicon. We could then design a different model, using different principles and rules, for English. To the extent that the two language-specific models differ from each other in a way that is inconsistent with what is known independently about language acquisition and linguistic universals, they would be observationally adequate but fall short of explanatory adequacy. An explanatory model would be one that correctly captures the distinction between the fixed universal parts and the parts that can change through learning, given the evidence of such processes, hence it would comprehend sentences in any language when supplied with the (1) fixed, universal components and (2) the information acquired from experience. We are interested in a theory of language comprehension that is explanatory in this sense.

Because speakers of different languages differ from each other, we need to use different *speaker models* to model the language processing and recognition grammars for different languages, and more generally for different speakers. A speaker model is a combination of the fixed, universal properties of human language capacities and individual properties acquired during language

comprehension. An explanatory adequate grammar will posit just enough fixed universal structure and individual variation to capture the properties of all languages. This could then be accompanied by a formal learning theory which creates speaker models from linguistic and other experiences.

Suppose we have constructed a theory of language comprehension that can be argued to be observationally adequate and explanatory. Does it also agree with data obtained from neuro- and psycholinguistic experimentation? Realistic language comprehension involves several features that an observationally adequate explanatory theory need not capture. One such property concerns automatization. Systematic use of language in everyday communication allows the human brain to automatize recognition and processing of linguistic stimuli in a way that an observationally adequate explanatory model might or might not be concerned with. Furthermore, real language processing is sensitive to top-down and contextual effects that can be ignored when constructing a theory of language comprehension. That being said, the amount of computational resources consumed by the model should be related in some meaningful way to reality. If, for example, the model engages in astronomical garden pathing when no native speaker exhibits such inefficiencies, the model can be said to be insufficient in its ability to mimic real language comprehension. We say that if the model's computational efficiency and performance behavior is in line with evidence from real speakers, it is *psycholinguistically adequate*. I will adopt this criterion as well. Psycholinguistic adequacy cannot be addressed realistically without addressing observational and explanatory adequacy first, but the vice versa is not true.

Language comprehension is viewed in this study as a species of cognitive information processing that processes information beginning from linguistic sensory input towards semantic interpretations. The framework closely resembles that of (Marr, 1982). Information processing will be modeled by utilizing *processing pathways*. The sensory input is conceived as a linear string of phonological words that may or may not be associated with prosodic features. Lower-level processes, such as those regulating attention or separating linguistic stimuli from other information such as music, facial expressions or background noise, are not considered. Because the input consists of phonological words, it is presupposed that word boundaries have been

worked out during lower-level processing. Since the input is represented as a linear string, we will also take it for granted that all phonological words are presented in an unambiguous linear order.

The output contains a set of semantic interpretations. The sentence *John saw the girl with a telescope* may mean that John saw a girl who was holding a telescope, or that John saw, by using the telescope that he himself had, the girl. This means that the original sensory input must be mapped to at least two semantic interpretations. These semantic interpretations must, furthermore, provide interpretations that agree with how native speakers interpret the input sentences.

The ultimate nature of semantic representations is a controversial issue. One way around this problem, adopted here, is to focus on selected aspects of semantic interpretation and try to predict them on the basis of the input. For example, we could decide to focus on the interpretation of thematic roles and require that the language comprehension model provides for each input sentence a list of outputs which determine which constituents appearing in the input sentence has which thematic roles. For example, we could require that the model provides that *John admires Mary* is processed so that John is judged the agent, Mary the patient, and not the other way around. The advantage is that we can predict semantic intuitions in a selective way without taking a strong stance towards the ultimate nature and implementation of the semantic notions. Another advantage is that we can focus on selected semantic properties without trying to understand how the semantic system works as a whole.

A third advantage of this approach is that we do not need to decide *a priori* what type of linguistic structures will be used in making these semantic predictions. We can leave that matter open for each theory to settle on its own way and require that correct semantic intuitions, in whichever way they are ultimately represented in the human brain, be predicted. Of course, any given model must ultimately specify how these predictions are generated. I will adopt what I consider to be the standard assumption in the present day generative theory and assume that input sentences are interpreted semantically on the basis of representations at a *syntax-semantics interface*. The

syntax-semantic interface is considered a level of representation (ultimately a collection of neuronal connections) at which all phonological, morphological and syntactic information processing has been completed and semantic processing takes over. It is also called Logical Form or *LF interface*. I will use both terms in this document. This means, then, that the language comprehension model will map each input sentence into a set of LF interface objects, which are interpreted semantically.⁵

It is not possible to construct a model of language comprehension without assuming that there occurs a point at which something that is processed from the sensory input is used to construct a semantic interpretation. A syntax-semantic interface seems to be a priori necessary on such grounds. Different theories make different claims regarding where they position that interface and what properties it has. It is possible to assume that the interface is located relatively close to the surface and operates with linguistic representations that have been generated from the sensory objects themselves by applying only a few operations. Meaning would then be read off from relatively shallow representations. An alternative, a “deep” theory of the syntax-semantic interface, is a theory in which the sensory input is subjected to considerable amount of processing before anything reaches this stage. We can build the model by assuming that there is only one syntax-semantic interface, or several, or that the linguistic information flows into that system all at once, or in several independent or semi-independent packages. The only way to compare these alternatives, and indeed many imaginable alternatives, is to examine to what extent they will generate correct semantic interpretations for a set of input sentences; it is pointless to try to use any other justification or argument either in favor or against any specific model or assumption. In

⁵ Almost any model can be interpreted in these terms. If we assumed that the input is something less than a linear string of phonological words, then the model must include lower-level information processing mechanisms that can preprocess such stimuli, perhaps ultimately the acoustic sounds, into a form that can be used to activate lexical items, in a specific order. We can also assume more, for example, by providing the model additional information concerning the input words, such as morphological decompositions, morphosyntactic structure or part-of-speech (POS) annotations.

general, then, we do not posit further restrictions or properties to the syntax-semantic interface apart from its existence.

3.2 Computational generative grammar framework

Scientific hypotheses must be justified by deducing the observed facts from the proposed hypothesis. The data, deductive calculations and the theories and hypotheses must be provided in a completely unambiguous form. The linear phase grammar and the material contained in this document, as well as all advanced natural sciences, are based on the same methodological principle (1).

(1) *Methodological principle*

Data, theories and the deductive chains connecting the two must be unambiguous.

Although the linear phrase grammar satisfies (1), it represents only one hypothesis or theory that does this. We say that a grammar or grammatical theory that satisfies (1) is a computational generative grammar (CGG). Thus, the linear phase grammar is one computational generative grammar among many other possible computational generative grammars. The alternative to (1) is a grammar or grammatical theory in which the theory, data and the deductive chains connecting the two are ambiguous and as such can be provided with several, often infinitely many, possible interpretations. This is obviously something we will want to avoid at all costs.

Let us consider then how to eliminate ambiguity from linguistics. A linguistic theory or grammar is a system that defines a set of attested (grammatical) expressions in some human language and only them. In addition to defining the grammatical expressions, we expect that it provides the grammatical expressions with some additional properties such as syntactic analyses, semantic interpretations, felicitous pragmatic contexts and other properties, depending on the scope of each individual study. Let us say that the surface expressions themselves (suppose they are linear sequences of phonological words) and their other properties are jointly called *expressions*. One way to construct a grammatical theory is to provide grammatical rules together with lexical items

which can be used to construct grammatical expressions in some language, namely by applying the grammatical rules to the lexical items and to their own output until well-formed expression is emerges. We call a grammatical theory of this type an *enumerative grammar*, since it defines the set of grammatical expressions by literally generating it. Another possibility – implemented by the linear phase grammar – is to formulate the grammatical theory in the form of a “filter” which takes expressions as input and then recognizes whether they are grammatical or not, and if they are, provides them with the further attributes we are looking for (semantic interpretation etc.). This type of theories can be called . However, despite the different ways these grammars work, both enumerative and recognition grammars are computational generative grammars: they define the set of attested expressions and provide them with further properties. They differ in that an enumerative grammar defines the set of enumeration, that is, enumerating its positive members, while the recognition grammar defies it by implementing its characteristic function, telling whether any arbitrary input expression belongs to the set or not.

Suppose, then, that we have constructed a computational generative grammar G that is either enumerative grammar or recognition grammar, and that we have some expression E in our dataset, E being either grammatical or ungrammatical. How can we show, assuming (1), that G entails correct properties of E ? Suppose that G is formulated as an enumerative grammar. If E is grammatical, we have to demonstrate that there is a chain of deductive calculations, call it *derivation*, which begins from the axioms of the theory and the relevant lexical items and ends up up with E . We can accomplish this by exploring recursively all possible derivations (from the relevant lexical set) until E is encountered. If all possible derivations have been exhausted while E has not been constructed, then the grammar is not observationally adequate. If E is ungrammatical, we do the opposite and establish, by using the same procedure, that there is no derivation that entails E . A computational generative grammar is a system consisting of the lexical items, grammar and the executive functions which implements the checking procedure above. The *dataset* will contain all expressions E, \dots that we think bear on the hypothesis we want to test. The general idea can be illustrated as follows:

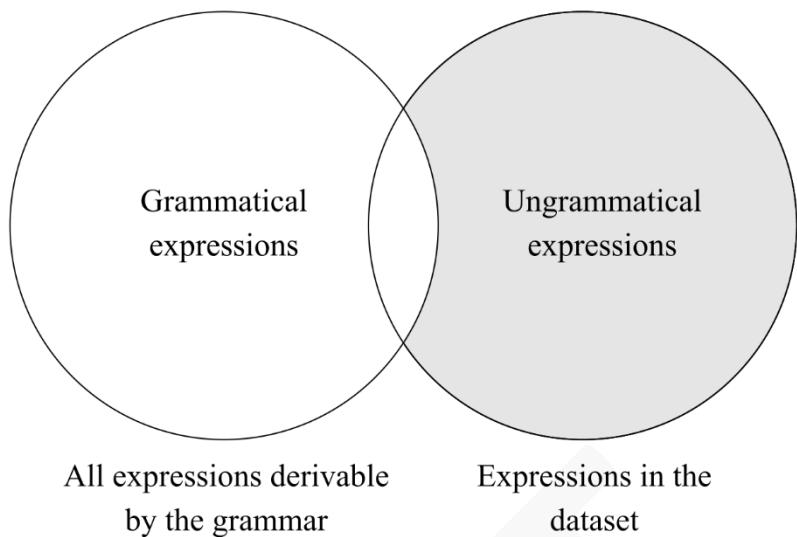


Figure 1. The computational generative grammar framework

The data and the dataset is unambiguous, since we have provided it in the form of an explicit dataset; the theory or hypothesis is unambiguous since it is provided in some machine-readable form; and the justification of the theory or hypothesis by the data is unambiguous, since every expression in the dataset is either linked with the theory/hypothesis by means of a completely explicit derivation or we have shown, by exhaustive search, that no derivation exists.

Recognition grammar implements a characteristic function of the set of attested (grammatical) expressions that may but does not have to generate expressions. We can construct a trivial recognition grammar by using an enumerative grammar as a platform: we dissolve the input expression E into its constituent lexical items and then examine whether E can or cannot be derived. Successful derivation will generate syntactic analyses and semantic interpretations for E . However, because the recognition grammar implements a characteristic function for the set of grammatical expressions, it does not have to enumerate all expressions that can be derived from the same words; it can recognize (and reconstruct) the properties of E by inspecting E alone. The system is therefore the same as the one provided in Figure 1, except that “all expressions derivable by the grammar” is replaced with “all expressions accepted by the grammar”.

There is a subtle but important difference between the enumerative and recognition grammars when it comes to practical work. An enumerative grammar will generate a set of grammatical expressions from the given lexical items, and is evaluated against data by making sure that unattested expressions are *not generated*. That is, anything not generated by the grammar is assumed to be ungrammatical. A recognition grammar must be tested explicitly against all relevant expressions, including ungrammatical expressions, and the force of the justification depends on the number and type of expressions there are in the dataset. It is possible, however, to use mere lexical selection in the testing of a recognition grammar as well: we test the algorithm against all linear permutations of the input words.⁶

Let us consider some practical aspects of this system. Let us say that one *study* refers to a dataset calculated by some particular CGG. We want to arrange things so that anybody can replicate the study without providing any further information or parameters. The script that runs the study must then be designed in such a way that it can be downloaded fully preconfigured and that it generates all required calculations for the selected dataset. We also want to arrange things so that a permanent record of these calculations will remain after the calculations have been done. This will be the function of *output files* that the CGG produces while it calculates the dataset. When a study is replicated, it should produce the same (or at the very least functionally equivalent) output files. The input and output files constitute the *raw input and output* of the study, since they ultimately explicate how the properties of the dataset follow(ed) logically from the axioms of the theory in some particular computational experiment.

⁶ This functionality is included in the linear phase model, although it has not been used in any published study to date. The researcher first creates a set of lexical items (numeration) which contains the relevant bound and free morphemes we want to test. The algorithm then generates a dataset by enumerating all possible linear permutations of those elements (while not violating the free/bound distinction) and tests the model against that dataset.

4 The core syntactic engine

4.1 Introduction

This section provides the minimal specifications that makes it possible to build up the kernel of the model (i.e. its syntactic engine) from scratch. The text is written for somebody working with a concrete algorithm or wants to understand the logic of the source code. Empirical matters dealing with the linguistic theory are discussed in the published literature and are omitted here.

4.2 Merge-1

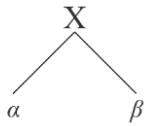
Linguistic input is received by the hearer in the form of sensory stimulus. We can first think of the input, to simplify the situation to bare essentials, as a one-dimensional string $\alpha \oplus \beta \oplus \dots \oplus \gamma$ of phonological words. In order to understand what the sentence means, the human parser (which is part of the human language faculty) must create a sequence of abstract syntactic interpretations for the input string received through the sensory systems. One fundamental concern is to recover the hierarchical relations between words. Let us assume that while the words are consumed from the input, the core recursive operation in language, Merge, arranges them into a hierarchical representation. For example, if the input consists of two words $\alpha \oplus \beta$, Merge yields $[\alpha, \beta](2)$.

(2) John \oplus sleeps.

\downarrow \downarrow

[John, sleeps]

The first line represents the sensory input consisting of a linear string of phonological words, and the second line shows how these words are put together. The same operation can be described in an image format as follows:



Here X denotes the complex constituent $[\alpha, \beta]$, α is the left constituent, β the right constituent.

What do we mean by saying that they are “put together”? While the two words in the sensory input are represented as two independent objects, once they are put together in syntax they are represented as being part of the same one unit, some type of linguistic chunk. We can attend to both of them as one object, manipulate them as part of the same representation, and in general perform operations that takes them both into account. Therefore, operation (2) presupposes that there exists a formally defined notion of phrase structure that is able to represent entities of this type.

Example (2) suggests that the syntactic component combines phonological words. While this is a possibility, it would not be linguistically useful assumption. We assume that (2) is mediated by *lexicon*: a storage of linguistic information that is activated on the basis of the original phonological words. The lexicon maps phonological words in the input into *lexical items* which contain *lexical features* § 4.9.4 such as lexical categories (noun, verb), inflectional features ('third person singular') and meaning ('John', 'sleeping'). We will assume from this point on that syntax is operating with lexical items, not with phonological words.⁷ It is possible, however, to omit the phonology-lexical mapping in its entirety and feed the model directly with the lexical items.

The resulting complex chunk $[\alpha, \beta]$ is asymmetric and has a *left (or first) constituent* and a *right (second) constituent*. The terms “left”/“first” and “right”/“second” are mnemonic labels and do not refer to concrete leftness or rightness at the level of neuronal implementation. Their purpose is to distinguish the two constituents from each other. They are related to leftness indirectly: since

⁷ Whether phonological information flows into syntax is an interesting empirical question. The current algorithm allows some phonological information to pass into further processing stages, but the empirical data motivating this assumption is quite peripheral.

we consume the sensory input from left to right, (2) implies that the constituent that arrives first will be the left constituent.⁸ Thus, we can think of the left constituent as the “first constituent.”

The theory itself places no limited on the number of elements that can be assembled together, although there are several other cognitive bottlenecks which can limit the resulting structures and our ability to process them. Suppose the next word is *furiously*. Example (3) shows three possible attachment sites, all which correspond to different hierarchical relations between the words.

- (3) a. [[John *furiously*] sleeps] b. [[John, sleeps] *furiously*] c. [John [sleeps *furiously*]]

The operation illustrated in (3) differs in a number of ways from what constitutes the standard theory of Merge at the time of present writing. I will label the operation in (3) by Merge-1, symbol -1 referring to the fact that we look the operation from an inverse perspective. Instead of generating linear sequences of words by applying Merge, we apply Merge-1 on the basis of a linear sequence of words in the sensory input and assemble the structure from left to right. The ultimate syntactic interpretation of the whole sentence is generated as a sequence of partial phrase structure representations, first [John], then [John, sleeps], then [John [sleeps, *furiously*]], and so on, until all words have been consumed from the input.

Several factors regulate Merge-1. One concern is that the operation may in principle create a representation that is ungrammatical and/or uninterpretable. Alternative (3)(a) can be ruled out on such grounds: *John furiously* is not an interpretable fragment. Another problem of this alternative is that it is not clear how the adverbial, if it were originally merged inside subject, could have ended up as the last word in the linear input. The default left-to-right depth-first linearization algorithm would produce **John furiously sleeps* from (3)a. Therefore, this alternative can be rejected on the grounds that the result is not consistent with the word order discovered from the input. If the default linearization algorithm proceeds recursively in a top-

⁸ The left to right assembly mechanism was proposed originally by (Phillips, 1996).

down left-right order, then each word must be merged to the *right edge* of the phrase structure, right edge referring to the top node and any of its right daughter node, granddaughter node, recursively. Under these assumptions a left-to-right model will translate into a grammatical model in which the grammatical structure is expanded at the right edge.⁹ This leaves (3)(b) and (c). The parser will select one of them. Which one? There are many situations in which the correct choice is not known or can be known but is unknown at the point when the word is consumed. An incremental parsing process must nevertheless make a decision. Let us assume that all legitimate merge sites are ordered and that these orderings generate a recursive search space that the algorithm will explore by backtracking. The situation after consuming the word *furiously* will thus look as follows, assuming an arbitrary ranking:

(4) *Ranking*

- a. [[John *furiously*], sleeps} (Eliminated, not consistent with linear order)
- b. [[John, sleeps] *furiously*] (Priority high)
- c. [John [sleeps *furiously*]] (Priority low)

The parser will merge *furiously* into a right-adverbial position (b) and branch off recursively. If this solution does not produce a legitimate output, it will return to the same point by backtracking and try solution (c). Every decision made during the parsing process is treated in the same way. All solutions constitute potential phrase structures, which are ordered in terms of the ranking. The mechanism can be illustrated with the help of a standard garden-path sentence such as (5).

(5)

- a. The horse raced past the barn.
- b. The horse raced past the barn fell.

⁹ These assumptions depend of course on prior assumptions concerning linearization. The present model assumes an exceptionless left-to-right depth-first algorithm. Suppose the researcher assumes that linearization is based on some alternative rule L; then Merge-1 must consider different alternatives based on which configurations could have in principle produced the input string relative to L.

Reading (5)b involves extra effort when compared to (a). At the point where the incremental parser encounters the last *for fell*, it has (under normal language use context) created a partial representation for the input in which *raced* is interpreted as the past tense finite verb, hence the assumed structure is [_{DP/S} *The horse*][_V *raced*][_{PP} *past the barn*] (DP/S = a subject argument, V = verb, PP = preposition phrase). Given this structure, there is no legitimate right edge position into which *fell* could be merged. Once all these solutions have been found impossible, the parser backtracks and considers if the situation could be improved by merging-1 *barn* into a different position, and so on, until it discovers a solution in which *raced* is interpreted as a noun-internal relative construction [_{DP} *The horse raced past the barn*]*fell*. This explanation presupposes that all solutions at each stage are ranked, so that they can be explored recursively in a well-defined order. Thus, at the stage at which *raced* is merged-1, the two solutions *raced* = finite verb and *raced* = participle are ranked so that the former is tried first.

Backtracking can be interpreted in one of two ways. If looked from the point of view of the recognition grammar, putting questions of performance aside, its purpose is to consider all possible derivations available given the input and some grammar. That is, the procedure will find all attested ambiguities from the input and rule out unattested syntactic and semantic analyses. Had we limited the search only to the first-pass parsing solution most of the derivations latently present in our grammar would not have been tested for correctness. Once we have an observationally adequate model which finds all attested ambiguities from the input and only those, we can interpret the search as a psycholinguistic process and try to map it with behavioral evidence acquired from psycholinguistic experiments. This is, however, optional. In addition, it is not necessarily true that speakers use full recursive backtracking in real language comprehension. If they do not use recursive backtracking, then that operation must be interpreted as a calculation device or testing procedure while additional mechanisms model realistic language comprehension. Aspects of realistic language comprehension are discussed in § 5.

Merge-1 can break constituency relations established at an earlier stage. This can be seen by looking at representations (2) and (3)c, repeated here as (6).

(6)	John	sleeps	furiously
	↓	↓	↓
[John,	sleeps]		

\rightarrow [John [sleeps furiously]]

During the first stage, words *John* and *sleeps* are sisters. If the adverb is merged as the sister of *sleeps*, this no longer holds: *John* is now paired with a complex constituent [*sleeps furiously*]. If a new word is merged with [*sleeps furiously*], the structural relationship between *John* and this constituent is diluted further, as shown in (7).

(7)	[John [[sleeps furiously] γ]
-----	------------------------------

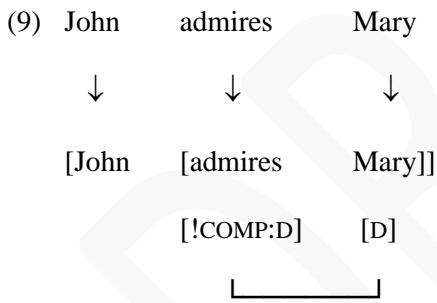
One consequence of this is that if we merge two words as sisters, we do not know if they will maintain the same or any close structural relationship in the derivation's future. In (7), they don't: future merge operations break up constituency relations established earlier. Consider the stage at which *John* is merged with a wrong verb form *sleep*. The result is a locally ungrammatical string **John sleep*. But because constituency relations can change in the derivation's future, we cannot rule this step locally as ungrammatical. It is possible that the verb 'sleep' ends up in a structural position in which it bears no meaningful linguistic relation with *John*. Only those configurations or phrase structure fragments can be checked for ungrammaticality that cannot be tampered in the derivation's future. Such fragments are called *phases* in the current linguistic theory.

4.3 Lexical selection features

Consider a transitive clause such as *John admires Mary* and how it might be derived under the framework described so far (8).

(8)	John	admires	Mary
	↓	↓	↓
[John	[admires Mary]]		

There is evidence that this derivation matches with the correct hierarchical relations between the three words. The verb and the direct object form a constituent that is merged with the subject. If we change the positions of the arguments, the interpretation is the opposite: Mary will be the one who admires John. But the fact that the verb *admire*, unlike *sleep*, can take a DP argument as its complement must be determined somewhere. If we do not determine it, then sentences such as **John admires to want him* or **John admires admires Mary* will be interpreted as grammatical. Let us assume that such facts are part of the lexical items: *admire*, but not *sleep*, has a lexical feature $[\text{!COMP:D}]$ which says that it requires a DP-complement. The fact that *admire* has the lexical feature $[\text{!COMP:D}]$ can be used by Merge-1 to create a ranking. When *Mary* is consumed, the operation checks if any given merge site allows/expects the operation. In the example (9), the test passes: the label of the selecting item matches with the label of the new word.



Feature $[\text{COMP:L}]$ means that the lexical item *licenses* a complement with type $[L]$, and $[\text{!COMP:L}]$ says that it *requires* a complement of the type $[L]$. Feature $[-\text{COMP:L}]$ says that the lexical item does *not* allow for a complement with label L .¹⁰ When the parser is trying to sort out an input, it uses lexical features (among other factors) to rank solutions. When the phrase structure has been completed, and there is no longer any input to be consumed, these features can be used for filtering. Filtering is performed at the LF interface (discussed later) and will be called the *LF*

¹⁰ The behavior of $[\text{!COMP:D}]$ and $[-\text{COMP:D}]$ is intuitively obvious, but there is redundancy between licensing and when there is no feature at all, as in the latter case the complement is neither mandated nor rejected, hence it could also be considered ‘licensed’. Licensing is currently used in the parsing process and also in some grammatical roles such as in distinguishing thematic heads from nonthematic ones; whether it could be eliminated in its entirety from the theory is an open question.

legibility test. It checks if the solution provided by the parser makes sense from a semantic point of view. Sentence **John admires to want him* will be judged ungrammatical because *admire* cannot select infinitival complement clauses.

Let us return to the example with *furiously*. What might be the lexical features that are associated with this item? There are three options in (9): (i) complement of *Mary*, (ii) right constituent of *admires Mary*, and (iii) the right constituent of the whole clause. We can rule out the first by assuming (again, for the sake of this example) that a proper name cannot take an adverbial complement. We are left with two options (10)a-b.

(10)

- a. [[s John [admires Mary]] furiously]
- b. [John [vp admires Mary] furiously]]

Independently of which solution is more plausible (or if they both are equally plausible), we can guide Merge-1 by providing the adverbial with a lexical selection feature which determines what type of left sisters it prefers, is allowed or required to have. I call such features *specifier selection features*. A feature [SPEC:S] (“select the whole clause”) favors solution (a), [SPEC:V] favors solution (b) (“select the verb phrase”). What constitutes a specifier selection feature will guide the selection of a possible left sister during comprehension.¹¹ Most of the lexical features are enumerated in § 4.9.4.

4.4 Phases and left branches

Let’s consider the derivation of (11).

¹¹ Because constituency relations may change at some later stage of the derivation, we do not know which pairs will constitute specifier-head relations in the final output. Therefore, we use specifier selection to guide the parser in selecting left sisters, and later use them at the syntax-semantics interface (LF interface) to verify that the output contains proper specifier-head relations.

(11)	John's	mother	admires	Mary.
	↓	↓	↓	↓
	[s[_{DP} John's mother] [VP admires Mary]]			

After the finite verb has been merged with the DP *John's mother*, no future operation can affect the internal structure of that DP. This is because Merge-1 is always to the right edge. All left branches therefore become *phases*: the derivation can forget them inside that particular parsing path.¹² It follows that if no future operation is able to affect a left branch inside the that parsing path, all grammatical operations (e.g., movement reconstruction) that must be done in order to derive a complete phrase structure must be done to each left branch before they are sealed off in this way. Furthermore, if after all operations have been done the left branch fragment still remains ungrammatical or uninterpretable, the original merge operation that created the left branch phase can be cancelled. This limits the set of possible merge sites. Any merge site that leads into an unrepairable left branch can be either filtered out as unusable or at the very least be ranked lower. Since the model can backtrack, it is able to reconsider all left branches. We can formulate the condition tentatively as (12).

(12) *Left branch phase condition*

Derive each left branch independently.

What we mean by deriving something “independently” becomes unambiguous only when we have a fully described grammar, but intuitively it says that if there are syntactic operations that must be done to the left branch before it is transferred to semantic interpretation and/or before we can consider further merge-1 operations, then such operations must be done at the point when the left branch is created. This is the intended interpretation of (12), although it leaves it open what these “operations” are.

¹² They are phases in the sense of (Chomsky, 2000, 2001), except that the units differ due to the reverse-engineered architecture. See (Brattico & Chesi, 2020).

4.5 Labeling

Suppose we reverse the arguments in (11) and derive (13).

(13) Mary admires John's mother

↓ ↓ ↓ ↓

Mary [admires [John's mother]]

Now recall that the verb's complement selection feature [$\neg \text{COMP:D}$] refers to the label [D] of the complement. What is the relationship between [D] and the phrase *John's mother* that occurs in the complement position of the verb in the above example? Since any constituent may contain an arbitrary number of elements, the mapping from complex constituents into labels like [D] is not trivial. How do we know, from any arbitrarily complex constituent, what its type is? The answer is provided by (14).

(14) *Labeling*

Suppose α is a complex phrase. Then

- a. if the left constituent is primitive, it will be the label; otherwise,
- b. if the right constituent is primitive, it will be the label; otherwise,
- c. if the right constituent is not an adjunct, apply (14) recursively to it; otherwise,
- d. apply (14) to the left constituent (whether adjunct or not).

The algorithm searches for the closest possible primitive head from the phrase starting from the top. “Closest” means closest from the point of view of the selector. If we analyzed the situation from the point of view of the labeled phrase itself, then closest is the highest or most dominant head. Principle (14) means that labeling ignores right adjuncts (this will be in fact a defining feature of adjuncts). Example (15) illustrates some basic examples.

- (15) a. [VP *admires Mary*] = left primitive verb;
b. [PP *from Mary*] = left primitive preposition;
c. [VP [DP *that man*][VP *admires Mary*]] = verb is the first left primitive;

- d. $[\text{DP } \text{the man}]$ = definite determiner is left, hence the label;
- e. $[\text{VP}[\text{VP } \text{admires Mary}]\langle\text{furiously}\rangle]$ = adjunct is ignored;
- f. $[\text{NP } [\text{DP } \text{that man's}] \text{ car}]$ = right primitive noun.

Notice that complex constituents $[\alpha \beta]$ themselves do not have labels or type; its label is determined by algorithm (14). The rule is defined in this way because this definition has generates correct and/or plausible results; the matter is empirical, however, and it is easy to experiment with different alternatives.¹³ Since the phrase structure geometry can change during incremental parsing, also labeling may change. This is shown by the following derivation:

- (16) a. $[\text{DP } \text{the man}] + \text{admires} =$
 b. $[\text{VP } [\text{DP } \text{the man}] \text{ admires}].$

What was originally a DP was transformed into a VP after the verb was merged. This is an important feature of parsing, because it means that we do not have to fit the incoming words into grammatical templates; the template is created dynamically as more words are assembled together.

The term *complex constituent* is defined as a constituent that has both the left and right constituent; if a constituent is not complex, it will be called *primitive constituent*.¹⁴ Consider again the derivation of (2), repeated here as (17).

- (17) John \oplus sleeps.

↓ ↓

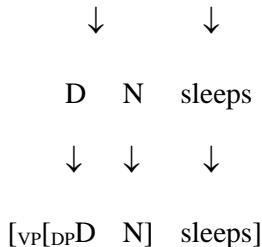
[John, sleeps]

¹³ For example, these assumptions rule out configuration $[\text{BP } \alpha^0 \beta]$ where a primitive left constituent does not project. We cannot therefore have a situation where a head moves to the specifier position of another head, in the terminology of standard linguistic theory. We can, however, contemplate modifications to the labeling algorithm such that this becomes a possibility.

¹⁴ A constituent that has only the left or right constituent, but not both, will be primitive according to this definition.

If *John* is a primitive constituent, labeling will provide [_{NP} *John sleeps*] which means that the sentence is interpreted as a noun phrase. However, (17) is a sentence or verb phrase, not a noun phrase. For example, it cannot be selected by transitive verbs (**John admires John sleeps*). We therefore assume that *John* is a complex constituent with the structure [_{DP} D N]. This information comes from the lexicon, where proper names are decomposed into D + N structures. The structure of (17) is therefore (18).

- (18) John ⊕ sleeps.



The structure of the lexicon is examined in detail in § 4.9.2, at this point it is important to note that we have nowhere assumed that phonological words are mapped into primitive lexical items. While some words are, others aren't. The point is, however, that whether some word is or isn't decomposed is crucial for labeling, because the labeling algorithm is sensitive to the distinction between primitive and complex constituent.

4.6 Upward paths and memory scanning

Constituents making up phrase structure representations can enter into several types of dependency relationships. Consider the following pair of Finnish sentences.

- (19) a. Pekka ei ihaillut *Merja-n/ Merja-a.
 Pekka not admire Merja-acc Merja-par
 'Pekka did not admire Merja.'
- b. Pekka on ihaillut Merja-n/ *Merja-a.
 Pekka has admired Merja-acc Merja-par
 'Pekka did admire Merja.'

The case form of the direct object argument depends on polarity: negative clauses require that the object is marked by the partitive case, glossed as PAR, while affirmative clauses require the accusative case, glossed as ACC. Both sentences, when generated by merge-1 from the input, generate a representation approximated in (20). The dependency between the negation and the direct object is shown by the line under the sentence.

(20) Pekka ei/on ihaillut Merja-a.

[Pekka [not/is [admire Merja]]]



This relationship cannot depend on local selection described in § 4.3. Dependencies of this type are modelled by relying on the notion of *path*. The case form at the case assignee must enter into a “compatibility check” with another element, the case assigner. In this case the case forms accusative and partitive enter into a compatibility check with the polarity element. It does this by forming an *upward path* from the case assignee to the case assigner. Suppose α is an element requiring checking; then

(21) *Upward path*

the upward path from α contains all constituents that dominate α plus their immediate daughters.

For example, if the phrase structure is (22)

(22) [_{AP} ... A [_{BP} ... B [_{α P} ... α ...]]]

then the upward path from α contains α P, BP and AP (the dominating constituents) plus their immediate constituents A and B. The dependency itself is formed by scanning through the path to a target element, and the first target element available is always selected. If the target is not found, the operation fails. This mechanism is applied to case checking, binding, adjunct positioning, control and operator scope calculations among many other nonlocal grammatical operations. The path coincides with or perhaps defines the content of the “syntactic working

memory” when α is processed.¹⁵ The path mechanism was originally inspired by the connectness approach of (Kayne, 1983, 1984).

4.7 Adjunct attachment

Adjuncts, such as adverbials and adjunct PPs, present a challenge for any incremental parser. Consider the data in (23).

(23) (Finnish)

- a. *Ilmeisesti* Pekka ihailee Merjaa.
 apparently Pekka admires Merja
 'Probably Pekka admires Merja.'
- b. Pekka *ilmeisesti* ihailee Merjaa.
 Pekka apparently admires Merja
- c. Pekka ihailee *ilmeisesti* Merjaa.
 Pekka admires apparently Merja
- d. ?Pekka ihailee Meja *ilmeisesti*.
 Pekka admires Merja apparently

The adverbial *ilmeisesti* ‘apparently’ can occur almost in any position in the clause. Consequently, it will be merged into different positions in each sentence. This confuses labeling and present a challenge for the parser, because the position of the adverb is unpredictable.

Adjuncts are assumed to be geometrical constituents of the phrase structure but they are stored inside a ‘parallel’ syntactic working memory making them invisible for sisterhood, labeling and selection in the primary working memory. They increase the dimensionality of the phrase structure. The idea is illustrated in Figure 24.

¹⁵ Recall that all left branches are phases, thus they have already been processed and moved out of the active syntactic working memory when α is targeted.

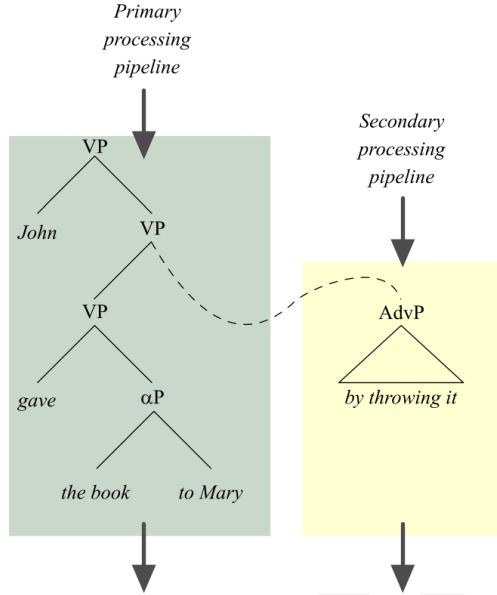


Figure 24. Adjuncts are geometrical constituents that are pulled out of the primary working memory and are processed inside separate processing pipeline. We can imagine that the VP is made up of two segments.

Thus, the labeling algorithm, as specified in § 4.5, ignores adjuncts. The label of (24) becomes V and not Adv: while analyzing the higher VP shell, the search algorithm does not enter inside the adverb phrase; the lower VP is penetrated instead.

- (24) [VP John [VP [VP admires Mary] <_{AdvP} furiously>]]

Consider (25) next.

- (25) John [sleeps <_{AdvP} furiously>]

The adverb constitutes the sister of the verbal head V and is potentially selected by it. This would often give wrong results. This is prevented by defining sisterhood in such a way that ignores right adjuncts. The adverb *furiously* resides in a parallel syntactic working memory and is only geometrically attached to the main structure; the main verb does not see it in the complement

position.¹⁶ The fact that adjuncts, like the adverbial *furiously* here, are optional follows from the fact that they are automatically excluded from selection and labelling. Whether they are present or absent has no consequences for either of these dependencies.

It follows from these assumptions, however, that adjuncts could be merged anywhere, which is not correct. Each adverbial (head) is associated with a feature linking it with a feature or set of features in its host structure. In this case the linking relation is established by a *tail-head dependency*: we imagine that the adjunct, which exists in the separate syntactic plane, constitutes a “hanging tail” related to the head. For example, a VP-adverbial is linked with V, a TP-adverbial is linked with T, a CP-adverbial is linked with C. Linking is established by checking that the adverbial (i) occurs inside the corresponding projection (e.g., VP, TP, CP) or (ii) can be linked with the corresponding head by means of an upward path. Conditions (i-ii) have slightly different content and are applied in different circumstances.¹⁷

(26) Condition on tail-head dependencies

A tail feature [F] of a head α can be checked if either (a) or (b) holds:

- a. α occurs inside a projection whose head β_F , or HP is β_F ’s sister;
- b. α can be linked with β_F through an upward path (21).

B_F denotes a head (or phrase) that has feature [F]. Condition (26)(a) is relatively uncontroversial. It states that a VP-adjunct must occur inside VP, or more generally, αP adjunct has to be inside a projection from α , where α is a feature. It does not restrict the position at which an adjunct must occur inside αP . Condition (26)(b) allows some adjuncts, such as preposition phrases, remain in a low right-adjoined or in “extraposed” positions in a canonical structure. If an adverbial/head

¹⁶ From the point of view of labeling, selection and sisterhood, the structure is therefore [VP[DP *John*] *sleeps*].

¹⁷ This is an imperfection in the model that needs resolving.

does not satisfy a tail feature, it will be reconstructed into a position in which it does during transfer. This operation will be discussed in § 4.8.6.

The definition of adjunction licenses both left and right adjuncts; what matters is that the adjunct is ignored by labeling and sisterhood. Left adjuncts are very similar to regular left phrases, however, because left phrases, whether adjuncts or not, are ignored by the labeling algorithm. Notice that the label of $[\alpha \beta]$ will be β if α is complex, it does not have to be an adjunct; if it is an adjunct, the situation will be the same. Left phrases and left adjuncts are not identical, of course, since the former can be selected while the latter can't; they generate different sisterhood dependencies. They differ also behaviorally. One difference is that left adjuncts can be stacked while left adjuncts cannot.

The definitions leave room for a situation where an adjunct phrase contains another adjunct phrase. We would then have the primary structure with a path into the secondary structure, and another path into the tertiary structure. This is of course perfectly normal in language where an adverbial phrase, for example, may contain further modifiers. There is no in principle limit on the number of adjunction dimensions any given structure can have, although they could become difficult to process.

4.8 Transfer

4.8.1 *Introduction*

After the spellout structure has been composed, it will be *transferred* to the syntax-semantics interface (LF interface) for evaluation and interpretation. Transfer performs a number of noise tolerance operations removing most or in some cases all language-specific properties from the input and deliver it in a format understood by the universal semantic system and the universal conceptual-intentional systems responsible for thinking, problem solving and other language-external cognitive operations that are part of the global cognition.

4.8.2 Phrasal reconstruction: general comments

A phrase or word can occur in a *canonical* or *noncanonical* position. A canonical position in the input could be defined as a position such that given the regular Merge-1 operations the constituent ends up in a position where it passes all LF interface tests. For example, regular referential or quantificational arguments must occur inside the verb phrase at the LF interface in order to receive thematic roles and satisfy selection properties of the verb. Example (27) illustrates a sentence where this is the case.

(27)	John	admires	Mary
	↓	↓	↓
[VP John [VP admires Mary]]			

Legitimate output is reached by merging-1 the input words into the phrase structure: the operation will bring all arguments inside the verb phrase where their correct thematic roles are determined.

Example (28) shows a variation in Finnish where this is no longer the case.

(28)	Ketä	Pekka	ihaille-e? (Finnish)
	who.par	Pekka.nom	admire-prs.3sg
'Who does Pekka admire?'			

The model generates the following first-pass parse for this sentence (in pseudo-English for simplicity):

(29)	[VP who [VP Pekka admires]]
------	-----------------------------

This solution violates two selection rules. There are two specifiers at the left edge, *who* and *Pekka*, and *admire* lacks a complement. Furthermore, even if the parser backtracks, it cannot find a legitimate solution. A candidate structure such as [VP [DP *who* *Pekka*] *admires*] is also

illegitimate.¹⁸ The two violations are not independent of each other, however: the element that triggers the double specifier violation at the left edge is the same element that is missing at the complement position. This is clearly not a coincidence: the interrogative pronoun causes these problems because it has been “dislocated” to the noncanonical position from its canonical complement position. We can see this by consider what would fix both problems: movement of the interrogative pronoun from the left edge to the complement position of the verb, thus something like *who Pekka admires* → *Pekka admires who*.

Dislocation is a general feature of all natural languages. In Finnish, for example, almost all word order variations of the referential arguments of a finite clause are possible. Application of the straightforward Merge-1 would result in a situation where each word order is represented by a different structure that do not necessary share any properties, yet most of these word orders only create stylistic differences. For example, the two orders in (30) correspond to the same core proposition ‘Pekka admires Merja’.

- (30) a. Pekka ihailee Merja-a. (SVO)
 Pekka.nom admires Merja-par
 'Pekka admires Merja.'
 b. Merja-a ihailee Pekka. (OVS)
 Merja-par admires Pekka.nom
 'Pekka admires Merja.'

Unless the system perform noise tolerance normalization where all selection restrictions can be checked, we cannot filter out ungrammatical sentences from the grammatical ones. It is clear, furthermore, that Finnish speakers use overt case forms, here the nominative and partitive, to do

¹⁸ In principle we could feed (29) directly to the LF interface component. This representation does not satisfy the complement selection features of the verb *admire*. If we do not control for what is filling the complement position of the verb, the model would accept **John admires* or even **John admires Mary to leave*. Similarly, the main verb is here preceded by two argument DPs, but this too can produce ungrammatical results such as **John Mary admires*.

this. The model must therefore “reverse-engineer” the construction in order to create a representation that can be interpreted at the LF interface. These operations take place during transfer.

Almost all reconstruction operations are based on the same computational template. First the system recognizes that the element occurs in an illegitimate position where it cannot satisfy one or several LF interface condition(s), such as selection or case checking. Once the system detects the presence of an offending constituent, it will attempt reconstruction, where a legitimate position is sought. If a legitimate position is found, the element is copied there, and the search ends; if a legitimate position is not found, the constituent remains at its surface position. In that case, some later operation may still dislocate it. How much variation is allowed in any given language depends on the properties of reconstruction and transfer: it is possible that transfer fails to recover the expression, resulting in deviance or ungrammaticality.

4.8.3 \bar{A} -chains

Let us return to the simple interrogative clause cited earlier, repeated in (31).

- (31) Ketä Pekka ihailee?
who.par Pekka.nom admires
[VP who [VP Pekka admires]]

Let us assume (to simplify) that Merge-1 creates the structure shown on the third line. The verb has two specifiers and no complement. One of these errors triggers reconstruction. In the current algorithm, it is triggered by the presence of an extra specifier *ketä* ‘who.par’. The reconstruction algorithm searches for the first gap for this element where it can generate a legitimate position and reconstructs (32).

- (32) [whe [Pekka [admires who]]]



This representation, which is created during transfer, is transferred to the LF-interface. The interrogative pronoun is copied from SpecVP into CompVP, and will be interpreted as the patient of the verb *admire*. Both selection violates have now been fixed: there are no longer two specifiers at the left end of the sentence, and the complement is no longer missing.

The search algorithm proceeds from the deviant element downstream until either a suitable position is found or there is no more structure. It cannot go upwards or sideways. In this case the operation begins from the sister of the targeted element and proceeds downward while ignoring left branches (phases) and all right adjuncts. This operation is called *minimal search* (§ 4.8.7). The element is copied to the first position in which (i) it can be selected, (ii) is not occupied by another legitimate element, and in which (iii) it does not violate any other condition for LF objects. If no such position is found, the element remains in the original position and may be targeted by later operation. If a position is found, it will be copied there; the original element will be tagged so that it will not be targeted second time.

This normalization operation is not yet sufficient. One problem is that the original sentence represents an interrogative clause – the speaker is asking rather than only stating something – that can only be selected by certain verbs. For example, while it is possible to say *John asked who John admires*, it is not possible to say **John claimed who John admires*. The bare VP representation does not yet represent this fact, because it was always interpreted (implicitly, as it were) as a declarative proposition. Interrogativization is represented by a *wh*-feature that must be part of the highest head in the sentence, so that it can be selected by *ask/claim*. This is accomplished by copying the *wh*-feature from the fronted interrogative pronoun to the local head (33)(a) or, if no local head is available, it is generated to the structure and then equipped with the *wh*-feature (b).¹⁹

¹⁹ Feature copying replaces feature checking of the enumerative generative grammar.

- (33) a. Who admires Pekka?
- [VP who_{wh} [VP admires_{wh} Pekka]]]
- b. who Pekka admires?
- who**_{wh} C_{wh} [Pekka [admires who_{wh}]]]

C_{wh} represents the extra head generated to the structure to carry the *wh*-feature. Notice that in both cases, a higher verb can now select the interrogative feature irrespective of whether it is inside the verb or the C-element (34).

- (34) a. John asks/claims + [VP_{+wh} who [admires_{V+wh} Pekka]]
- b. John asks/claims + [CP_{+wh} **who** C_{wh} [Pekka [admires who]]]]

The reconstruction algorithm generates the C head into the structure. We can think of this operation as reconstructing a phonologically null head, i.e. a head that had no direct representation in the sensorimotoric input.

The dislocated phrase also represents the *propositional scope* of the question. The interpretation is one in which the speaker is asking for the identity of the object of admiration ('which x: Pekka admires x'). Let us consider how dislocation represents the propositional scope of an interrogative clause. First, the reconstructed interrogative pronoun or phrase is called an *operator*. This means that an argument such as 'who' does not refer to anything by itself; it is an "argument placeholder." We can also think of it as a variable, the target of the question that must be filled in by the "answer." Then, any finite element that contains the same operator feature, in this case *wh*-feature, will determine the scope.

- (35) **who**_{wh} C_{wh,fin} does John admire who_{wh}



The operator-scope dependency, marked by the line in the representation above, is checked by an operator-variable module inside narrow semantics, which pairs the operator with its scope marker. If feature [OP] (e.g. *wh*) represents the operator, then the closest head with [FIN][OP] will be the

scope marker. Together they create the Fregean interpretation ‘which x: John admires x?’ . These dependencies, which have both a syntactic and semantic dimension, are called *Ā-dependencies* or *Ā-chains* in the literature. Chains have a syntactic side, because the dislocated deviant element must be reconstructed by a syntactic operation that takes place during transfer, by minimal search; and a semantic side, because the operator must be linked with the corresponding scope marked inside the semantic system. Both operations can fail, and when one or both fail, the input sentence is judged ungrammatical.²⁰²¹

Traditional enumerative grammars describe the same process – Ā-chains – by moving the interrogative pronoun from the complement position into the left edge position where it represents the scope of the question. The operation is called Move or Internal Merge. Reconstruction is its mirror image, a reverse-engineered Move operation; they describe the same object, a directionless chain that is constituted by the two (or more) members in two (or more) syntactic positions.²²

4.8.4 A-chains

In addition to Ā-chains, discussed in the previous section, many languages exhibit another type of phrasal reconstruction, creating A-chains. To see how, consider that English preverbal subject position can be occupied by both the agent and patient (36).

²⁰ This architecture was originally inspired by Chomsky’s dual interpretation model (Chomsky, 2008). Operator features involved in these mechanisms are interpreted by a special semantic system (the operator-variable module inside narrow semantics).

²¹ There is some variation with respect to how operator-variable constructions such as interrogatives are packaged for communication. In some languages, the operator remains in its canonical position and is not fronted at all to the beginning of the clause; then there are languages where several operators can be fronted. Thus, the fact that both Finnish and English front interrogative pronouns is not inevitable or necessary.

²² Whether chains are visible and, if they are, whether there is an operation that is sensitive to such objects are interesting nontrivial empirical questions. What we mean by using the term “chain” here is that that there is an operation (Move, Reconstruction) that brings them into existence. The operation linking variables to their scopes is not sensitive to the chain as such, but is based on lexical features + upward pathing and reads information created by reconstruction. However, there have been several versions of the algorithm in which chains were considered real objects on their own right when required by empirical evidence.

- (36) a. John admires Mary.

[John [admires Mary]]

- b. Mary was admired (by John).

[Mary [was [admired (by John)]]]

The direct Merge-1 solution (second lines in the above example) is wrong in the case of (b), because *Mary* would be in the agent position, *John* the patient (if present). However, we can conclude on the basis of data of this type (a–b) that the English preverbal subject position is *not* a thematic position. It can be occupied by both agents (a) and patients (b). This is supported further by the data from Finnish, which shows that argument order does not correlate necessarily with thematic interpretation. This brings us to the presence of the tensed auxiliary verb *was* in the example (b). Note that both the bare finite verb (a) and the aux-verb construction (b) involve a tensed finite verb; when the sentence contains an auxiliary, tense information is expressed by the auxiliary. Let us assume that tense represents an independent packet of grammatical information in the sentence that may be expressed either by combining it with the verb (a) or by carrying it by an otherwise dummy auxiliary verb (b). We can depict the situation by using the schema (37).

- (37) [_{TP} John [_{TP} T [_{VP} admire Mary.]]]

Tense T can be expressed either by an auxiliary (*was*, *does*) or by combining it with the main verb; the latter operation will be discussed later. Recall that we already assumed in §4.5 that phonological words such as *John* decompose into several lexical items D and N; now we apply this scheme to tensed verbs, and assume that they are decomposed into T and V. Having assuming this, we can express the intuition that the preverbal subject position is not a thematic position and that the thematic roles are assigned inside the VP: the preverbal subject position is SpecTP and the thematic agent position SpecVP. Then we can assume that the grammatical subject is reconstructed from the former into the latter (38).

- (38) [_{TP} John₁ [_{TP} T [_{VP} John₁ [_{VP} admire Mary.]]]

This operation called *A-reconstruction* and it forms an *A-chain*. The operation is triggered by the fact that SpecTP is not a thematic position – hence the referential argument cannot remain at this position at the LF-interface – and the operation locates the first position where it can be interpreted. The fact that SpecTP is a nonthematic position is marked in the current implementation by feature [EF]. A-reconstruction differs from \bar{A} -reconstruction in that the former does not form operator-variable constructions inside the semantic system and is not triggered by operator features such as *wh*-features.²³

Representation (38) is very close to how the model represents canonical transitive or intransitive clauses. We assume that they are composed out of at least two separate lexical items T and V, which together form a *complex predicate*. Thematic roles are assigned inside the VP, while T is an extra part of the complex predicate that does not assign a new thematic role but modifies the predicate itself. That is, the sentence means ‘John T + V Mary’, where T and V are part of the same predicate. There is only one detail that is still missing, namely the difference between transitive and intransitive verbs.

4.8.5 Head reconstruction

Let us consider again (39).

(39) John admires Mary.

So far we have assumed that the tensed finite verb is made of at least two grammatical heads or independent packages of grammatical information – tense T and the verb stem V. The consequence is that the sentence has one extra position that is nonthematic, namely the preverbal subject position SpecTP. We noted that the T can be expressed either by an independent auxiliary element (*does*, *was*) or by the finite verb, but what was left unexplained was how the finite verb

²³ The EPP no longer exists as a monolithic lexical feature but is replaced by a function EF that relies on a set of edge features. When these details do not matter I will keep using the term “EPP.”

is dissolved into the two independent components when that strategy is used. *Head reconstruction* solves this issue.

First we should note that whether and how grammatical heads are chunked into morphological and phonological words is to some extent arbitrary and subject to crosslinguistic variation. Thus in English, we can express the same tense information by *John does admire Mary* and *John admires Mary*, where in the latter the two heads have been chunked together. Head reconstruction must therefore contain an operation that recovers morphological chunks from phonological words. This part is handled by the lexico-morphological module. It takes phonological words as input and decomposes them into their constituent morphemes, thus *admires* ~ T + V. These heads are forwarded into the syntactic component and merged-1 to the syntactic structure.

In the first stage, a *complex head* T(V)⁰ is created inside the syntactic component. Thus, the lexical streaming operation maps lexical decompositions into complex heads: *admires* ~ T + V ~ T(V)⁰. *Head reconstruction* extracts the heads and distributes them into the structure, T(V)⁰ ~ [T ... [...V...]]. The steps are illustrated in (40).

- (40) John admires Mary. (Input)
[John [T(V)⁰ Mary]] (Output of Merge-1)
[John [T [V Mary]]] (Output of head reconstruction)

When V is extracted out from T, the closest possible reconstructed position is searched for and the element is merged into that position. The configuration shown in (40) is selected because T can select a VP complement.

We assumed earlier that the phrase structure is made up of primitive constituents, which are elements that have zero daughters, and complex constituents, which have two daughters. Thus, if α and β are primitive constituents, then $[\alpha \beta]$ is a legitimate complex constituent. Because $[\alpha \beta]$ will always be treated like a phrase, we can also say that it is a *phrasal constituent*. What we have not considered is a situation where the complex constituent has only one daughter constituent (left

or right). This configuration creates complex constituents that are nonphrasal, which, then, are the complex heads referred to above. Thus we have $\alpha(\beta)^0 = [\alpha \beta]$, where α is a constituent that has only one (right) constituent β . The mechanism is iterative, and allows the system to chain several heads into ever more complicated words, for example $A(B(C))^0 = [A [B C]]$. Phrasal syntactic operations do not treat nonphrasal complex constituents (complex heads) as phrases (observing the lexical integrity principle), while they can still contain several grammatical heads (primitive lexical items) ordered into a linear sequence.²⁴

These assumptions raise the question of why we compose complex heads at all instead of mapping the different parts of words directly into the syntactic structure. That is, if the lexico-morphological component can decompose *admires* into T + V, why not take these lexical items and merge-1 them directly in [...[T...[...V...]]]? In fact, earlier versions of the model did exactly that. The reason this was rejected later is because there is empirical evidence suggesting that grammar can create and manipulate complex heads. For example, in Finnish it is possible to move complex predicates as shown in (41).

- (41) Myy-dä-kö₁ Pekka halusi __₁ koko omaisuutensa?
- sell-a/inf-q Pekka wanted all possessions

‘Was it selling that Pekka wanted to do with all his possessions?’

The current model handles constructions like (41) exhibit long head movement by reconstructing complex heads syntactically. If we assume, then, that syntax has access to complex heads, why can’t we assume that such complex heads are already stored as complex heads in the lexicon? According to this alternative, the phonological word *admire* is mapped directly into $T(V)^0$, which is placed into syntax as a complex package. One reason why the algorithm does not work in this way is the existence of morphological parsing that is able to create and comprehend novel words

²⁴ Morphologically complex words are linear sequences of (sensorimotoric) elements. We can think of the lopsided constituency relation as representing simple “followed by” relation: produce α followed by β , in short $\alpha * \beta$. Each element in the sequence must be atomic and map into one sensorimotoric program.

that do not exist in the lexicon. This means that it must be possible for speakers to perceive words as composed out of individual morphemes. In other words, there must exist a process which can decompose phonological words into their constituent parts that do not correspond to anything in the lexicon. But there is another reason complex heads are not stored in the lexicon, namely this would mean that lexical items are ultimately syntactic objects and, by the same token, that the component storing and manipulating them is part of syntax. In the current theory, the lexico-morphological component is extrasyntactic and only understands linear sequences of patterns; once we have syntactic objects, we are in the domain of syntax proper. The distinction between lexico-morphological processes ('non-syntax') and syntax is just made in this way. Positing a third mediating structure, a syntactic lexicon, is possible but serves no purpose as far as the current evidence goes.²⁵

4.8.6 *Adjunct reconstruction (also scrambling reconstruction)*

Thus far we have examined \bar{A} -reconstruction, A-reconstruction and head reconstruction. There is a fourth reconstruction type called *adjunct reconstruction* (also adjunct floating, scrambling reconstruction). Consider the pair of expressions in (42) and their canonical derivations.

(42)

- a. Pekka käski meidän ihailla Merjaa.
 Pekka.nom asked we.gen to.admire Merja.par
 [Pekka [asked [we [to.admire Merja]]]]
 'Pekka asked us to admire Merja.'
- b. Merjaa käski meidän ihailla Pekka.
 Merja.par asked we.gen to.admire Pekka.nom

²⁵ There are grammatical framework and theories which posit complex syntactic lexicons in this exact sense. Whether they are better than the present model must be tested with real data in computational experiments. Likewise, the hypothesis that we can do by mapping linear sequences of morphemes directly into syntactic structure, as assumed currently, is an empirical hypothesis.

[Merja [asked [we [to.admire Pekka]]]]

'Pekka asked us to admire Merja.'

Derivation (b) is incorrect. Native speakers interpreted the thematic roles as being identical in these examples. The subject and object are in wrong positions. Neither Ā- nor A-reconstruction can handle these cases. The problem is created by the grammatical subject *Pekka* ‘Pekka.nom’, which has to move upwards/leftward in order to reach the canonical LF-position SpecVP. Because the distribution of thematic arguments is very similar to the distribution of adverbials in Finnish, one can argue that richly case marked thematic arguments can be promoted into adjuncts. Case forms are morphological reflexes of tail-head features. If the condition is not checked by the position of an argument in the input, then the argument is treated as an adjunct and reconstruction into a position in which the condition is satisfied. In this way, the inverted subject and object can find their ways to the canonical LF-positions (43).

(43) [$\langle \text{Merjaa} \rangle_2$ T/fin [___₁ [käski [meidän [ihailla [___₂ [$\langle \text{Pekka} \rangle_1$]]]]]]]

Merja.par asked we.gen to.admire Pekka

'Pekka asked us to admire Merja.'

Notice that because the grammatical subject *Pekka* is promoted into adjunct, it no longer constitutes the complement; the partitive-marked direct object does.

4.8.7 Minimal search

Let's consider again a standard interrogative sentence such as (44).

(44) Who does [_A John's brother] admire → [_B every day]?

We have assumed that the interrogative pronoun is reconstructed into the canonical complement position marked by the gap ___. The fronted interrogative and the gap form an \bar{A} -chain. What has been left unsaid is how this reconstruction actually happens.

All reconstruction is based on *minimal search*. The gap position is located by descending downwards through the phrase structure from the position of the reconstructed element and detecting the first legitimate position (hence the term “minimal”). More formally, at each step the algorithm assumes a phrase structure γ as input and moves to α in $\gamma = \{\alpha P \alpha \beta\}$ unless α is primitive, in which case β is targeted.²⁶ The operation therefore moves downwards on the phrase structure by following selection and labelling. Left and right (adjunct) phrases are ignored. The operation never branches since the algorithm provides a unique solution for any constituent. Consider (44). To reach the reconstruction site $__1$, the search algorithm must avoid going into the subject A and into the temporal adverbial B. This is prevented by minimal search, which follows the projectional spine of the sentence and ignores both left specifiers/adjuncts and right adjuncts (there are no “right specifiers” in the model).²⁷

4.8.8 *AgreeLF*

Most languages exhibit an agreement phenomenon, in which some element, such as finite verb, agrees with another element, typically the grammatical subject. This is illustrated in (45).

- (45) a. John admires Mary
- b. *John admire Mary
- c. *They admires Mary

The third person features of the subject are cross-referenced in the finite verb. The term agreement or phi-agreement refers to a phenomenon where the gender, number or person features (collectively called phi-features in this document) of one element, typically a DP, covary with the same features on another element, typically a verb, as in (45).

²⁶ Notation $\{\alpha \beta\}$ refers to $[\alpha \beta]$ or $[\beta \alpha]$.

²⁷ There is a possible deeper motivation for minimal search. Both left branches and right adjuncts are transferred independently as phases and are therefore no longer in the current syntactic working memory. The notion of minimal search coincides with the contents of the current syntactic working memory at any point in the derivation.

A typical argument DP like *Mary* has interpretable and lexical phi-features which are connected to the manner it refers to something in the real or imagined world. Thus, *Mary* refers to a ‘third person singular individual’. A predicate, on the other hand, must be linked with an argument that has interpretable phi-features. To model this asymmetry, a predicate will have *unvalued* phi-features, denoted by $[\varphi_]$. The value will be usually provided by the argument with which the predicate agrees with. This called *valuation*. The starting point is (46).

(46) Mary	[admires	John]
‘argument’	‘predicate’	
D N	T/fin	
↓	↓	
[phi:num:sg]	[phi:num:_]	
[phi:per:3]	[phi:per:_]	
[phi:det:def]	[phi:det:_]	

The operation that fills the unvalued slots for the predicate by the phi-features of the argument is *AgreeLF* (47).

(47) Mary	[admires	John]
[phi:num:sg]	[phi:num:sg]	
[phi:per:3]	[phi:per:3]	
[phi:det:def]	[phi:det:det]	
└─ AgreeLF ─┘		

As a consequence, the unvalued features disappear from the lexical item and the predicate is now linked with its argument.

Some predicates arrive to the language comprehension system with overt agreement information. The third person suffix *+s* signals the fact that the argument slot of *admires* should be (or is)

linked with a third person argument. In many languages with sufficiently rich agreement, overt phrasal argument can be ignored. Example (48) comes from Italian.

(48) Adoro Luisa.

admire.1sg Luisa

‘I admire Luisa.’

Thus, the inflectional phi-features are not ignored; instead, they are extracted from the input and embedded as morphosyntactic features to the corresponding lexical items, as shown in (49).

(49) John admire + s Mary.

↓ ↓ ↓

[John [admire Mary]]

[...3SG...]

This means that *admires* will have both unvalued phi-features and valued phi-features as it arrives to syntax from the sensory input and the lexico-morphological system. While this might be considered unintuitive, the former exists due to the fact that *admires* is lexically a predicate, while the latter are extracted from the input string as inflectional features (50).

- (50) a. admire- = lexical predicate, hence it has unvalued phi-features [\emptyset _];
b. -s = third person singular valued inflectional phi-features [\emptyset].
c. = [\emptyset _] + [\emptyset].

If the head has valued phi-features, then AgreeLF must also check that there is no phi-feature conflict. Thus, a sentence such as **Mary admire John* will be recognized as ungrammatical.

4.8.9 Ordering of operations

Ā/A-reconstruction and adjunct reconstruction presuppose head reconstruction, because heads and their lexical features guide Ā/A- and adjunct reconstruction. The former relies on EPP features and empty positions, whereas the latter relies on the presence of functional heads.

Furthermore, Ā/A-reconstruction relies on adjunct reconstruction: empty positions cannot be recognized as such unless orphan constituents that might be hiding somewhere are first returned to their canonical positions. The whole transfer sequence is therefore (51).

(51) *Ordering of operations during transfer*

- a. Reconstruct heads →
- b. Reconstruct adjuncts →
- c. Reconstruct A/Ā-movement →
- d. Agree-1 →
- e. LF-interface and legibility →
- f. Semantic interpretation.

The sequence is performed in a reflex-like manner. It is not possible to evaluate the operation only partially or to backtrack.²⁸

4.9 Lexicon and morphology

4.9.1 *From phonology to syntax*

Most phonological words enter the system as polymorphemic units. They are made of stems, derivational bound morphemes, inflectional affixes, clitics, and prosodic information. The *lexico-morphological component* performs a mapping from input words into grammatically meaningful units by utilizing the *lexicon*. The lexicon is a list of *lexical entries* which are all key-value pairs where the key is a symbol that can occur in the input, and the value specifies the type of item it is

²⁸ The sequence provided in (51) is compatible with two possible implementations, one where all operations are executed during one cycle and another where they form several cycles. In an ideal case, the algorithm would travel through the structure only once and fix all errors, in the order of (51), as they are encountered. This corresponds to the first option. This unification has not been accomplished in the present version due to the fact that the adjunct reconstruction algorithm (b) still differs too much from (a), (b) and (c). In addition, the transfer algorithm triggers separate extraposition operations that fixes certain issues with adjuncts. It is clear that the adjunct handling is not optimal and, once this issue has been sorted out, the whole sequence (51) could possibly be executed inside just one cycle.

mapped to (currently primitive lexical item, morphologically complex item, inflection or clitic).

The lexicon is specified in ordinary text files the user can modify.

Let us consider the processing of morphologically complex words first. The lexicon matches phonological words with a *morphological decomposition*, if the latter is present. A morphological decomposition consists of a linear string of morphemes $m_1 \# \dots \# m_n$. These morphemes, which (we assume) designate primitive lexical items, retrieve the corresponding primitive lexical items $M_1 + \dots + M_n$ from the same lexicon which are fed into syntax, where they form complex heads $M_1(M_2 \dots (M_n))^0$. The symbols $M_1 \dots M_n$ are keys in the same lexicon. We can imagine the morphologically complex words as “morphological chunks,” lexicon information (e.g., $M_1 \dots M_n$) that has been packaged into one unit. Head reconstruction (§ 4.8.5) extracts them into head spreads $[M_1 \dots [M_2 \dots [\dots M_n \dots]]]$ (52).

- (52) a. John admires Mary. (Input)
b. John T+V Mary (Output of lexicon)
c. John T(V) Mary (Input to syntax)
d. [John [T [V Mary]]] (Output of head reconstruction)

Primitive lexical items are feature sets $\{f_1 \dots f_n\}$. It is also possible to map a phonological word directly into a primitive lexical item. For example, the English definite article *the* is currently modelled in this way (i.e. $\text{the} \sim D^0 \sim \{f_1 \dots f_n\}$).

Phonological strings mapping into primitive lexical items and strings mapping into complex decompositions exist in the same lexicon with no principled difference between the two. They are both stored in a dictionary structure with the phonological string as the key, and the feature set or morphological decomposition as value. It is possible to feed the algorithm with any string that exist in the lexicon. Thus, input strong $John \oplus T \oplus \text{admire} \oplus \text{Mary}$ generates in many cases the same outcome as *John admires Mary*, the only difference begin that steps (52)a–c are now omitted. Symbol T must, of course, exist in the lexicon as a key and map into a primitive lexical item containing tense information and other features that belong to this item. It is also possible to

feed the model directly with morphological decompositions, which skips the step (52)a. In this way, the researcher can test the model with (possible or impossible, attested or unattested) words that do not exist in the lexicon. Thus, input *John admire#s Mary* will generate the same output as *John admires Mary*, assuming that the phonological word *admires* maps to *admire#s* in the same lexicon. In this way, it is possible to simulate morphological parsing. We can also test the model with words that are impossible or otherwise unattested, for example, what would happen if English adpositions were combined with Finnish-style possessive suffixes, or what happens if we combine English *that* with the third person singular agreement affix (e.g., *that#s*).²⁹

Consider the following figure which contains a screenshot of a small segment from the file defining the content lexicon, specifically, the some of the entries for the verb ‘paint’.

```

206 maalatessa :: maalaan-#v#ESSA/inf LANG:FI
207 maalattua :: maalaan-#v#TUA/inf LANG:FI
208 maalaavan :: maalaan-#v#VA/inf LANG:FI
209 maalata :: maalaan-#v#A/inf LANG:FI
210 maalasi :: maalaan-#v#T/fin#[V] LANG:FI
211 maalaan :: maalaan-#v#T/fin LANG:FI
212 maalaan- :: PF:maalaan- LF:paint V !COMP:∅ SPEC:∅ CLASS/TR LANG:FI

```

The first six items represent complex phonological words (keys in the lexicon dictionary) which map into morphological decompositions, strings of further entries in the same lexicon. These entries all begin with *maalaan-* which maps into a primitive lexical item, the verb ‘to paint’ (line 212). This is the root stem. Its features are listed on the line 212, among them its overt phonological shape (PF:*maalaan-*), idiosyncratic meaning (LF:*paint*), major lexical category (V) and then several selection features.³⁰ The decompositions are then followed by one or several morphemes all separate from each other by # , representing morpheme boundaries (symbol = represents clitic boundary). Symbol v denotes a grammatical head associated with transitivity and voice, symbols ESSA/inf and A/inf are infinitival heads; they all exist in the same lexicon as

²⁹ We can bypass the content lexicon and feed the model with “general” lexical items standing for the major lexical categories, that is by strings such as D N T V D N (corresponding to *John admires Mary*) where D, N, T and V map to general determiner, noun, tense and verb units that have no specific content such as ‘John’ or ‘admire’.

³⁰ Features are ultimately processed as “formal patterns” and are represented as strings.

separate entries. Symbol [V] represents vowel lengthening and maps into the finite third person singular agreement morpheme in Finnish. These decompositions should match with the way the words are glossed or, stated in other way, they *are* the glosses provided in a format which the algorithm can interpret.

It is possible to have the same phonological string to map into two or more lexical entries. This creates ambiguous words, such as *bank*. When the model confronts an ambiguous word, all its lexical entries create their own parsing subtrees. This can be avoided, if it is not required for the research purposes at hand, by disambiguating the word manually in the input string. We could create, for example, two words *bank1* and *bank2* mapping into separate lexical entries and then use these symbols in the input. This makes the interpretation of the results clearer and easier when the irrelevant lexical entries are not contaminating the output.

The lexico-morphological component does not have access to the notion of complex constituent. It operates with linear objects of symbols matched and retrieved from the lexicon and which are then processed as linear objects. The lexical objects are fed into syntax as a linear stream, and only form complex constituents when syntax assembles them into complex heads and constituents.

4.9.2 *Lexical items and lexical redundancy rules*

Primitive lexical items that are stored in the lexicon (a dictionary data structure) are sets of features, which emerge from three distinct sources. One source is the language-specific lexicon, which stores information specific to lexical items in a particular language. For example, the Finnish sentential negation behaves like an auxiliary, agrees in φ -features, and occurs above the finite tense node in Finnish. Its properties differ from the English negation *not*. Some of these properties are so idiosyncratic that they must be part of the language-specific lexicon.

Another source of lexical features comes from *lexical redundancy rules*. For example, the fact that transitive verbs can select object arguments need not be listed separately in connection with each transitive verb. This pattern is provided by redundancy rules which are stored in the form of

feature implications ' $F_1 \dots F_n \rightarrow G_1 \dots G_m$ ' stating that if a lexical item has features ' $F_1 \dots F_n$ ' it will also get features ' $G_1 \dots G_m$ '. If there is a conflict between what is specified in the language-specific lexicon and in the redundancy rules, the language-specific lexicon wins. That is, the redundancy rules define what we can consider to be the default template or perhaps prototypical lexical items that are still subject to possible exceptions specified in the language-specific lexicon.

A repertoire of universal morphemes constitutes a third source. It contains elements like T, v and C. These are assumed to be present in all or almost all languages. Their properties too are modified by lexical redundancy rules.

One of the features in the language-specific lexicon linked with any item is its *language*. The feature is currently provided in the form [LANG:XX] where XX is a symbol for the language (e.g., FI = Finnish, EN = English, IT = Italian, HU = Hungarian and so on). Because language is represented as a feature, it can enter into lexical redundancy rules. In this way, it is possible to create *language-specific lexical redundancy rules* and, by using such rules, language-specific variations of common functional lexical items such as C, T or v.

Let us consider these assumptions by using an example, adpositions. Adpositions are represented by the general major lexical feature P, which is associated with a set of universal features by a lexical redundancy rule. For example, no adposition takes finite clause complements. Because these rules are specified as lexical redundancy rules, there can still be exceptions. Thus, if in some language there exist a special group of adpositions which take finite clause complements, it is possible override the lexical redundancy rule by providing the required selectional features in the language-specific lexicon.

Let us consider a more realistic example. While most adpositions require a noun phrase complement (*He ran towards me*, **He ran towards*), there are exceptions such as *He lives near (me)* that can be specified in connection with the individual lexical items. In this case, then, there is a general pattern that is subject to a few individual exceptions. Finnish adpositions however differ from English adpositions in that the former, but not the latter, can exhibit agreement and

occur in postpositional form (*lähellä minua* ‘near I.par’, *minun lähellä-ni* ‘I.gen near-px/1sg’). This is something that can be modeled by relying on language-specific lexical redundancy rules of the form ‘preposition + LANG:FI → agreement + postpositional configuration’ which add the required features (whatever they are) into the Finnish adpositions. If such properties are only associated with a group of adpositions, then it is possible to form an additional feature representing the relevant grouping (much like we have intransitive and transitive verb stems) and add that feature to the redundancy rule.

4.9.3 Derivational and inflectional morphemes

Inflectional and derivational morphemes differ in how they are processed. Derivational morphemes are processed as described above: they are mapped into primitive lexical items, i.e. feature sets, and are streamed into syntax where they form complex heads and then head spreads via head reconstruction (§ 4.8.5). Inflectional features, too, are mapped into feature sets, but these features are inserted *inside* an adjacent lexical item in word and, as a consequence, in the lexical stream. For example, the verb *admires* is decomposed into three elements *admire#T#3sg* by morphological decomposition in the lexicon, where the last element represents the third person agreement features. This is symbol mapped into the corresponding feature set, phi-features (singular, third person) in this case, which are inserted inside T as lexical features, thus *admire#T#3sg* ~ $T_{3\text{sg}}(V)^0$ ~ [...[$T_{3\text{sg}}$ [...V...]]]. If we specified in the lexicon that 3sg refers to a derivational element, then a separate agreement head such as Agr^0 would be generated. Sometimes there is evidence that agreement projects separate heads rather than just features, sometimes the issue is unclear (as in the case of Finnish possessive suffixes). Both hypotheses can be tested easily as just explained.

4.9.4 Lexical features

The catalog of lexical features undergoes changes as the model is being applied to new empirical phenomena. Some of the most important lexical features are explained in this section. Most of them can only be interpreted against the empirical data they were originally created to capture.

All features are subject to the *feature conflict rule* which bans any derivation that contains a feature [F] and its counter feature [–F] at the LF-interface.

[PF:X] provides the phonological matrix of the element, currently just a string. [LF:X] provides its idiosyncratic semantic interpretation. [COMP:L], [!COMP:L] and [–COMP:L] license, mandate or block complements of the type L, where L can be any feature inside the head of the complement but usually refers to one of the major lexical categories . The relevant notion of complement of α is the sister of α with the exclusion of $[\beta^0 \alpha]$ where β is primitive. The head is determined by labeling algorithm § 4.5. [()COMP:*] applies the rule to any head, thus [!COMP:] designates a functional head that has some complement. [SPEC:L], [!SPEC:L] and [–SPEC:L] license, mandate or block specifiers of the type L, which prevents any non-adjunct constituent with a head containing feature L to occur at the edge of α . [TAIL:L] requires that the element tails a head with feature L. [TAIL:L₁,...,L_n] requires a match with a set of features [TAIL:L₁,...,L_n].

4.10 Narrow semantics

4.10.1 Syntax, semantics and cognition

Semantics is the study of meaning. In this study we construct the notion of meaning in the following way. We assume that linguistic conversation and/or communication projects a set of semantic objects that represents the things that the ongoing conversation or communication “is about.” This set or structure contains things like persons, actions, thoughts or propositions as represented by the hearer and the speaker. This temporary semantic repository is called *global discourse inventory*. The discourse inventory can be accessed by global cognition, operations like thinking, decision making, planning, problem solving and others. Linguistic expressions, in turn, are utilized to introduce, remove and update entities in the discourse inventory. Consider the sentence *the horse raced past the barn*. We assume that the hearer projects the following entities into the global discourse inventory as s/he processes the sentence.

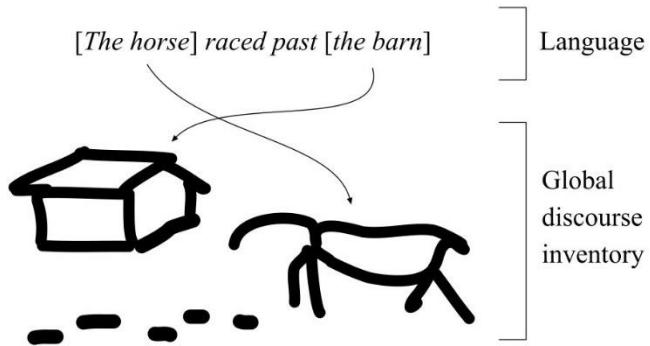


Figure 37. Narrow semantics projects semantic objects into the global discourse inventory.

We assume that this process of projecting and updating entities inside the global discourse inventory gives linguistic expressions their “meaning” in the narrow technical sense relevant to the present work.

The hypothesis that linguistic expressions provide a vehicle for updating the contents of the discourse inventory requires that there exists some mechanism which translates linguistic signals into changes inside the global discourse inventory. This mechanism, or rather collection of mechanisms, is called *narrow semantics*. It can be viewed as a gateway that encapsulates the syntactic pathway and mediates communication in and out. Cognitive systems that are outside of narrow semantics belong to global cognition and incorporate language-external cognitive processes. Narrow semantics is implemented by special-purpose functions or modules which interpret linguistic features arriving through the syntax-semantics interface. We can perhaps imagine language together with the narrow semantics as a cognitive system that grammaticalizes extralinguistic cognitive resources.

Different grammatical features are interpreted by qualitatively different semantic systems. Information structure (notions such as topic and focus) is created by different processing pathway than the system that interprets quantifier scope. Narrow semantics can therefore be also viewed as a gateway or “router,” where the processing of different features is distributed to different

language-external systems or where the language faculty or syntactic processing pathway makes contact with other cognitive systems (see Figure 38).

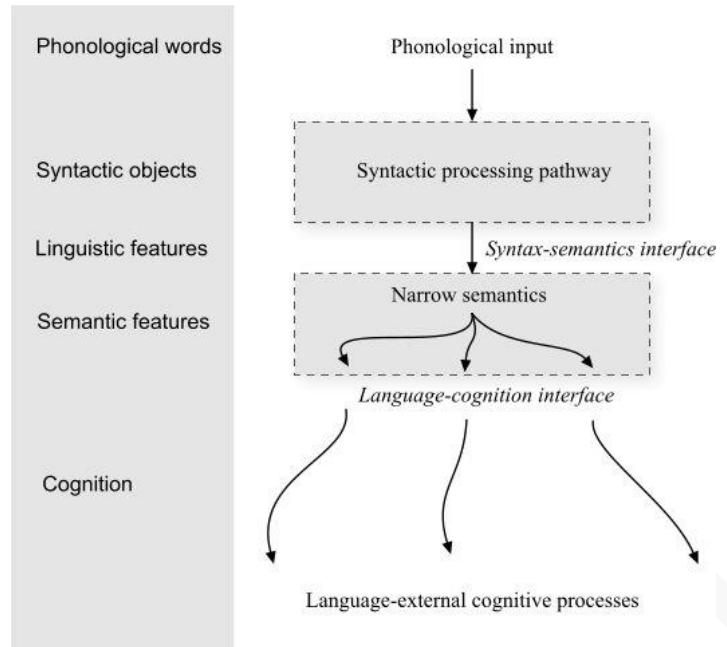


Figure 38. The general cognitive architecture. Narrow semantics bleeds input from language and feeds it in a transformed shape to other cognitive systems with which it makes contact. What these connections are or can depend on the innate neuronal architecture of the human brain.

A proper name such as *John* can be thought of as referring to a simple “thing” like an individual person. A pronoun like *he*, on the other hand, can refer in principle to several objects in the discourse inventory (*John₁ admires Simon₂, and he_{1,2} is very clever*), and the same is true of quantifiers like *every man* or *two men*. Furthermore, expressions like *no one* do not refer to anything, yet they, too, have meaning. Even *John* can be ambiguous if there are several men with the same name. The general problem is that there is nothing in the expression itself that determines unambiguously what it denotes, so the listener must perform some type of selection and/or guessing. To handle this, all expressions are first linked to unambiguous semantic entries inside narrow semantics, representing their intrinsic semantic properties, and these intermediate representations are transformed into actual denotations that point into semantic objects in the

global discourse inventory accessed by other cognitive processes. To illustrate, consider a short conversation *The horse raced past the barn; it was very fast*. The first sentence will establish that there are two things in the global discourse inventory the sentence speaks about: the horse and the barn. The next sentence then makes a claim about some “it” that we must link with something. This pronoun can denote four things in this conversation: the horse, the barn, the whole event, or a third entity not yet mentioned, as shown in Figure 39.

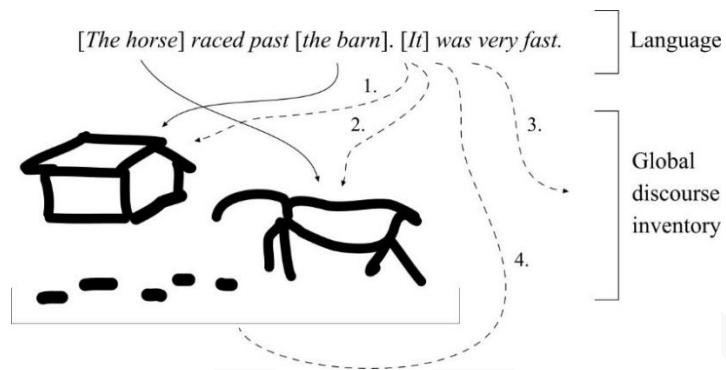


Figure 39. Possible denotations for expression *it*.

Narrow semantics calculates possible denotations by using the properties of the referring expression itself (e.g., nonhuman, singular, third party in the conversation) and what is contained in the global discourse inventory at the time when the expression is interpreted (the horse, the barn, the event, a possible third entity). The truth-value of the sentence, and thus its ultimate meaning in a context, is calculated when all referential expressions are provided some denotation. This is called *assignment*. The most plausible assignment in this case is one in which *the horse* denotes the horse, *the barn* denotes the barn, and *it* denotes the horse as well. The assignment in which *it* denotes the barn is implausible, but possible in principle. The model provides all possible assignments and their rankings as output. Assignments are calculated at the language-cognition interface.

When a linguistic feature such as [SG] ‘singular’ is transformed into a formal understood by global cognition, what we mean is that the formal signal representing that feature in the linguistic output will activate a corresponding signal or representation (representing, say, ‘one’) inside the

extralinguistic cognitive system. In the case of quantifiers such as *some*, the mapping is more complex but the principle is the same. This quantifier signals that we are supposed to select some (but no matter what and how many) objects from the global discourse inventory. I assume that the operation of ‘selecting some’ is part of the human cognitive repertoire accessed by narrow semantics, and that this is the reason a quantifier (or a lexical feature corresponding to it) can exist.

4.10.2 Argument structure

Argument structure refers to the way referential arguments are organized syntactically and semantically around their predicates. Let us consider some of the fundamental properties of this system. Consider the situation depicted in Figure 8.

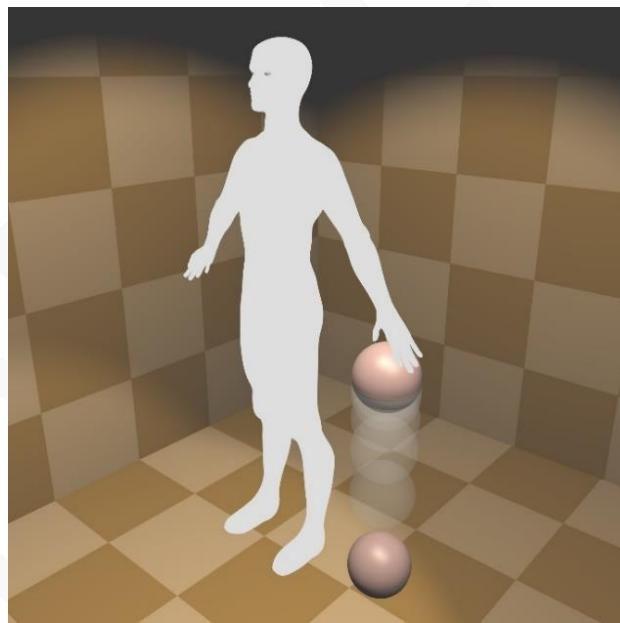


Figure 8. A non-discursive, perceptual or imagined representation of the meaning of the sentence *John dropped the ball*.

Imagine you have a non-discursive (analogous, continuous) experience of a person dropping a ball. We assume that instead of a static figure, you are looking at a dynamic scene that evolves over time. The visual-conceptual experience will be structured in such a way that we perceive the situation as involving a spatiotemporally continuous person, the ball and an event where the

person drops the ball and the ball falls as a consequence. Finally, the event ends when the ball touches the floor. These are the entities that would exist, or come into existence, in the discourse inventory at the moment we experience the event. The grammatical elements that make up the sentence *John dropped the ball* grammaticalize these conceptual objects, in this case John, the ball and the event of dropping expressed by the verb. Part of the meaning emerges from the structure into which they have been embedded. Lexical items and their structure are linked with the contents of the discourse inventory and, ultimately, the everyday human experience. Let us consider the grammatical representation of the sentence (53) and how the different parts may correlate with the conceptual experience depicted in Figure 8.

(53) [_{TP} John₁ [_{TP} T_{pst} [_{vP} ____₁ [_{vP} v(drop) [_{VP} fall [the ball]]]]]]]

The combination of the object the ball and the verb, the lower VP, correspond to a subevent where the ball falls. Thus, a bare verb together with an object is typically interpreted as denoting an event that involve one participant. This is combined with the small verb v and the agent argument John, which is interpreted as the causer of the subevent. That is, John is the agent of the event because he causes the ball to fall. The combination of *fall* + v will be the transitive verb *drop* ‘cause to fall’. Consequently, the argument at SpecvP is interpreted as the agent of the event, while the argument inside the VP is interpreted as the patient that must undergo the event caused by the agent. These interpretations are created when the representation arriving at the LF-interface is interpreted semantically by narrow semantics. Tense T adds tense information to the event, but the argument at SpecTP will not get a secondary thematic interpretation. In sum, the lexical items and their surrounding structures map into entities in the discourse inventory and human experience, in this case objects, events, causation and temporal properties. The elements must be organized in some specific way in order for this interpretation to succeed, and the formal selection restriction features present in the lexicon guide the syntactic processing pathway towards

solutions which satisfy these ultimately semantic requirements grounded in the structure of human experience.³¹

4.10.3 Predicates, arguments and antecedents

Predicates are characterized by the fact that they must be linked with arguments (§ 4.8.8). A predicate is defined as such by unvalued phi-features $\varphi_{_}$, representing an “unsaturated argument” in the Fregean sense. The same property can be represented by feature [ARG] suggesting “needs an argument” which under the current model automatically generates an uninterpretable phi-set $\varphi_{_}$ to the lexical item.

A predicate α is linked to an argument at the LF-interface/narrow semantics by finding first and closest argument from its *edge*, where edge refers to α 's (i) own feature content (sublexical pro-elements), (ii) complement, (iii) specifier(s) and finally to (iv) any constituent within the upward path (21) § 4.6. The search space defined by (i-iv) is explored in this order, and the first potential argument found will be selected. That is, if the head contains a sufficiently rich pro-element (e.g., in virtue of strong agreement), that will represent the argument (54)a; if not, then if the head has a referential complement, it is selected (54)b; if not, then the specifier is examined (54)c; if no suitable referential specifier exists, then the whole upward path is explored (54)d–e.

- (54) a. Lähde-n kotiin.
leave-prs-1sg home

‘I leave home’

- b. Kohti talo-a.

³¹ Categories like ‘causing the ball to fall’ or even ‘the ball’ are not dichotomous categories one could easily ‘program’ in Python. What happens, instead, is that the speakers and hearers project these categories to the world and/or to their own experiences. John is conceptualized as the cause of the event because we perceive, correctly or incorrectly, that he initiates the action. Yet merely willing the ball to fall is not sufficient; John must also perform actions that allow the gravity to take over. If an external agency controls John’s mind, then the sentence *John dropped the ball* would be false if *John* refers to the person, but true if we use *John* to refer to his physical body, as in the sentence *the tree dropped its leaves*. In this sense sentences provide “perspectives” to reality by conceptualizing parts of it.

- towards house-par
 ‘towards the house’
- c. Sitä päätti [Pekka (taas) [lähteä aikaisin]].
 expl decide-pst-3sg Pekka again leave early
 ‘Pekka decided to leave early again.’
- d. Pekka sanoi että __ halua-a lähteä.
 Pekka said that want-prs.3sg leave
 ‘Pekka said that he wants to leave.’
- e. Pekka wants to sleep.

Failure to find an argument from αP results in a situation where the argument is searched from a nonlocal domain (54)d–e, capturing control.

The argument-predicate mechanism interacts with AgreeLF. When the argument is located *inside* the complement, it is invisible for the predicate-argument mapping rule. AgreeLF detects it and copies its phi-features to the head, which in some cases is sufficient to interpret the argument (rule i). In some cases this is not sufficient, or it does not occur at all, and the argument must be literally copied to the specifier position of α in order to be visible for α (rule iii). This captures the EPP phenomenon in the current model and creates A-chains § 4.8.4.³²

4.10.4 The pragmatic pathway

In addition to the syntactic processing pathway, there exists a separate pragmatic pathway that monitors the incoming linguistic information and uses it to create pragmatic interpretations for the illocutionary act and/or communicative situations associated with the input sentence. The same pragmatic pathway computes topic and focus properties to the extent that they have not been

³² The idea of capturing the EPP/Agree complex by relying on a theory of predication comes from by (Rothstein, 1983) and has been developed by many others, for example (Chomsky, 1986, p. 116; Heycock, 1991; Kiss, 2002; Rosengren, 2002; Williams, 1980). The notion of edge that plays a key role in the present model is developed on the basis of (Chomsky, 2008).

grammaticalized and/or are based on general psychological characteristics of the communicative situation. This means that what linguistics describe as ‘information structure’ is partitioned into an interpretative extralinguistic pragmatic component and a syntactic (grammaticalized) component.

The pragmatic pathway works by allocating attentional resources to the incoming expressions and the corresponding semantic objects and by notifying if an element is attended in an unexpected (‘too early’ versus ‘too late’) position. The current implementation relies on two syntax-pragmatics interfaces to handle these cases. The first interface occurs very early in the processing pipeline – at the lexical stream currently – and registers all incoming referential expressions and allocates attentional resources to them. Another interface is connected to the component implementing discourse-configurational word order variations during transfer. It responds to situations in which some expression occurs in a noncanonical ‘too early’ or ‘too late’ position, and then generates the corresponding pragmatic interpretations ‘more topical’ and ‘more focus-like’, respectively. The results are shown in the field “information structure” in the output. By positioning the expression into a certain noncanonical position the speaker wanted specifically to control the attentional resource allocation of the hearer, and the hearer then infers that this must be the case.

Properties of the pragmatic pathway can also grammaticalize into *discourse features* that are interpreted by the pragmatic pathway. Discourse features are marked by [DIS:F]. Thus, it is possible for language also to mark topics or focus elements by using special features (which may also be prosodic). The idea that notions generated by language-external cognitive systems grammaticalize inside the syntactic pathway is one of the core principles of the semantic system.

4.10.5 Operators

Operators are elements that create operator-variable dependencies, usually (perhaps exclusively) represented by \bar{A} -chains § 4.8.2, 4.8.3. Interrogative clauses (55) represent a prototypical example of operators and operator-variable structures.

(55) What did John found __?

Operators are processed inside an operator-variable module by linking lexical elements containing operator features with a scope-marker that determines the propositional scope for that particular operator. The operation is triggered when narrow semantics sends a lexical element containing an operator feature such as [wh] for the operator-variable module for interpretation. The scope marker is closest lexical head inside the upward path with [wh][FIN]. The finite operator cluster [wh][FIN] is created during reconstruction. The relevant configuration is illustrated by (56).

(56) What did_{wh,fin} John found (what)?

Here the operator feature at the in situ interrogative pronoun *what* triggers the binding mechanism, which locates the scope marking element at C having features [wh] and [FIN] and generates the meaning ‘what x: John found x’.

4.10.6 *Binding*

All referential expressions such as pronouns, anaphors or proper names must be linked with some object or objects in the discourse inventory so that both the hearer and speaker know what is “talked about.” This is a nontrivial problem for language comprehension, because all such expressions are ambiguous. Even a proper name such as *John* can refer to any male person in the current conversation whose name is or is assumed to be John. The model restricts possible denotations by using whatever lexical features are available in the lexical items themselves and whatever is available in the global discourse inventory. For example, a proper name *John* can only denote a single male person what that name. In the same way, *she* cannot denote a male person. Both can denote something that does not (yet) exist in the global discourse inventory. Some referential expressions also impose structure-dependent restrictions on what they can denote. Reflexives like *himself* must be coreferential with a nonlocal antecedent (57), *him* cannot denote a too local antecedent (58), and R-expressions such as proper names must remain free (59).

- (57) a. John₁ admires himself_{1,*2}.
b. *John's₁ sister admires himself₁

- (58) a. John₁ admires him_{*1,2}
b. John's₁ sister admires him_{1,2}

- (59) a. He₁ admires John_{*1,2}.
b. His_{1,2} sister hates John₁.

Binding theory is concerned with the conditions that regulate these properties. Since assignments are computed at the language-cognition interface, whatever mechanism drives binding must regulate what takes place there. This is implemented technically by assuming that nominal expressions (e.g., anaphora, pronouns, R-expressions) contain features that provide “instructions” for a cognitive system that then filters possible assignments, as shown in the data above.

5 Performance

5.1 Human and engineering parsers

The linear phase theory is a model of human language comprehension. Its behavior and internal operation should not be inconsistent with what is known independently concerning human behavior from psycholinguistic and neurolinguistic studies and, when it is, such inconsistencies must be regarded as defects that should not be ignored, or judged irrelevant for linguistic theorizing. In this section, I will examine the neurocognitive principles behind the model, their implementation, and also examine them in the light of some experimental data.

5.2 Mapping between the algorithm and brain

Figure 32 maps the components of the model into their approximate locations in the brain on the basis of neuroimaging and neurolinguistic data. The image is modified from (Brattico, 2023b).

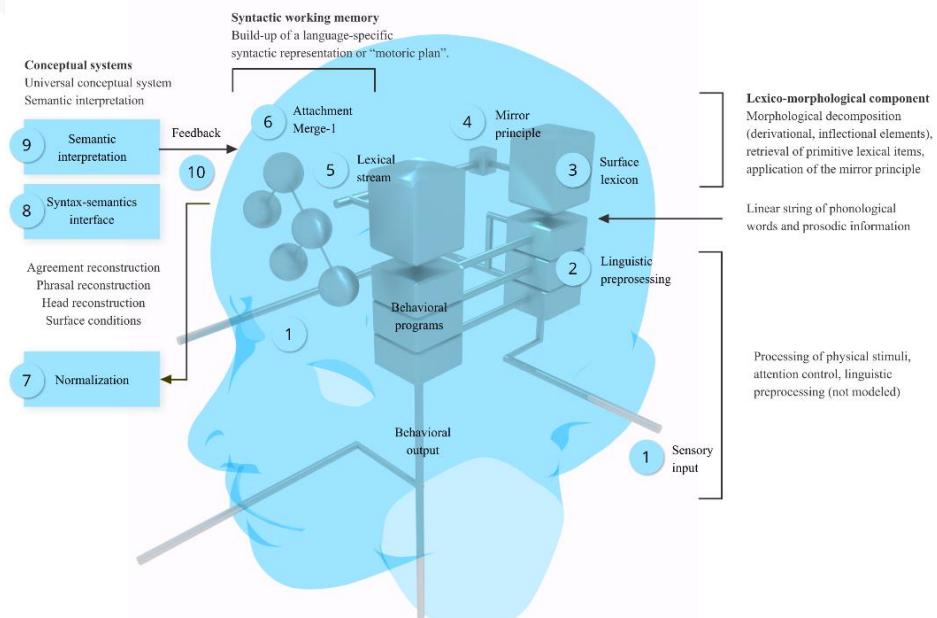


Figure 32. Components of the model and their approximate locations in the human brain. See the main text for explanation.

Sensory stimulus (1) is processed through multiple layers of lower-level systems responsible for attention control and modality-specific filtering, in which the linguistic stimuli is separated from other modalities and background noise, localized into a source, and ultimately presented as a linear string of phonological words (2). The current model assumes a tokenized string (2) as input. Brain imaging suggests that the processing of auditory linguistic material takes place in and around the superior temporal gyrus (STG), with further processing activating a posterior gradient towards the Wernicke's area that seems responsible for activating lexical items (3)(§ 4.3). Preprocessing is done in lower-level sensory systems that rely on the various modules within the brain stem. Activated lexical items are streamed into syntax (5). Construction of the first syntactic representation for the incoming stimulus is assumed to take place in the more anterior parts of the dominant hemisphere, possibly in and around Broca's region and the anterior sections of STG (6)(§ 4.1). It is possible that these same regions implement transfer (§ 4.8), as damage to Broca's region seem to affect transformational aspects of language comprehension. The syntax-semantic interface (LF-interface) is therefore quite conceivably also implemented within the anterior regions and can be assumed to represent the endpoint of linguistic processing. There is very little neurolinguistic data on what happens after that point. The architecture assumes that the endpoint of syntactic processing makes contact with other cognitive systems, possibly anywhere in the brain, via narrow semantics.

5.3 Cognitive parsing principles

5.3.1 *Incrementality*

The linear phase algorithm is incremental: each incoming word is attached to the existing partial phrase structure as soon as it is encountered in the input. Each word is encountered by the parser as part of a well-defined linear sequence. No word is put into a temporal working memory to be attached later, and no word is examined before other words that come before it in the linearly

organized sensory input. Apart from certain rearrangement performed by transfer normalization, no element that has been attached to the partial phrase structure being developed at any given point can be extracted from it later.

There is evidence that the human language comprehension system is incremental. It is possible to trick the system into making a wrong decision on the basis of incomplete local information, which shows that the parser does not wait for the appearance of further words before making its decisions. This can be seen from (60), in which the parser interprets the word *raced* as a finite verb despite the fact that the last word of the sentence cannot be integrated into the resulting structure.

(60)

- a. The horse raced past the barn.
- b. The horse raced past the barn fell.

The linear phase algorithm behaves in the same way: it interprets *raced* locally as a finite verb and then ends up in a dead end when processing the last word *fell*, backtracks, and consumes additional cognitive resources before it finds the correct analysis. The following shows how the algorithm derives (b). Step (11) contains the first dead end and starts the backtracking phase.

```

1   the + horse
2   [the horse]
3   [[the horse] T]
4   [[the horse] T] + race
5   [[the horse] T(V)]
6   [[the horse] T(V)] + past
7   [[the horse] [T(V) past]]
8   [[the horse] [T(V) past]] + the
9   [[the horse] [T(V) [past the]]]
10  [[the horse] [T(V) [past the]]] + barn
11  [[the horse] [T(V) [past [the barn]]]]
12  [[the horse] [T(V) [past [the barn]]] + T]
13  [[the horse] [T(V) [past [the barn]]]] + fell
14  [[[the horse]:11 [T [__:11 [race [past [the barn]]]]] T]
    [[[the horse]:11 [T [__:11 [race [past [the barn]]]]]] T] + fell
15  [[[the horse]:12 [T [__:12 race]]] past]
16  [[[the horse]:12 [T [__:12 race]]] past] + the
17  [[[the horse]:12 [T [__:12 race]]] [past the]]
18  [[[the horse]:12 [T [__:12 race]]] [past the]] + barn
19  [[[the horse]:12 [T [__:12 race]]] [past [the barn]]]
20  [[[the horse]:12 [T [__:12 race]]] [past [the barn]]] + T

```

```

15 [[[the horse]:12 [T [__:12 race]]] [past [[the barn] T]]]
[[[the horse]:12 [T [__:12 race]]] [past [[the barn] T]]] + fell
16 [[[the horse]:12 [T [__:12 race]]] [past [the [barn T]]]]
[[[the horse]:12 [T [__:12 race]]] [past [the [barn T]]]] + fell
17 [[[the horse]:12 [T [__:12 race]]] [[past [the barn]] T]]
[[[the horse]:12 [T [__:12 race]]] [[past [the barn]] T]] + fell
18 [[[the horse]:12 [T [__:12 race]]] <past [the barn]> T]
[[[the horse]:12 [T [__:12 race]]] <past [the barn]> T] + fell
19 [the [horse T]]
[the [horse T]] + race
20 [the [horse T(V)]]
[the [horse T(V)]] + past
21 [the [horse [T(V) past]]]
[the [horse [T(V) past]]] + the
22 [the [horse [T(V) [past the]]]]
[the [horse [T(V) [past the]]]] + barn
23 [the [horse [T(V) [past [the barn]]]]]
[the [horse [T(V) [past [the barn]]]]] + T
24 [the [horse [T(V) [past [[the barn] T]]]]]
[the [horse [T(V) [past [[the barn] T]]]]] + fell
25 [the [horse [T(V) [past [the [barn T]]]]]]
[the [horse [T(V) [past [the [barn T]]]]]] + fell
26 [the [horse [T(V) [[past [the barn]] T]]]]
[the [horse [T(V) [[past [the barn]] T]]]] + fell
27 [[the [horse [T [race [past [the barn]]]]]] T]
[[the [horse [T [race [past [the barn]]]]]] T] + fell
28 [[the [horse [T [race [past the]]]]] barn]
[[the [horse [T [race [past the]]]]] barn] + T
29 [[the [horse [T [race [past the]]]] [barn T]]
[[the [horse [T [race [past the]]]] [barn T]]] + fell
30 [[the [horse [T [race past]]] the]
[[the [horse [T [race past]]] the] + barn
31 [[the [horse [T [race past]]] [the barn]]
[[the [horse [T [race past]]] [the barn]]] + T
32 [[the [horse [T [race past]]] [[the barn] T]]
[[the [horse [T [race past]]] [[the barn] T]]] + fell
33 [[the [horse [T [race past]]] [the [barn T]]]
[[the [horse [T [race past]]] [the [barn T]]]] + fell
34 [[the [horse [T race]]] past]
[[the [horse [T race]]] past] + the
35 [[the [horse [T race]]] [past the]]
[[the [horse [T race]]] [past the]] + barn
36 [[the [horse [T race]]] [past [the barn]]]
[[the [horse [T race]]] [past [the barn]]] + T
37 [[the [horse [T race]]] [past [[the barn] T]]]
[[the [horse [T race]]] [past [[the barn] T]]]] + fell
38 [[the [horse [T race]]] [past [the [barn T]]]]
[[the [horse [T race]]] [past [the [barn T]]]] + fell
39 [[the [horse [T race]]] [[past [the barn] T]]]
[[the [horse [T race]]] [[past [the barn] T]]] + fell
40 [the horse]
[the horse] + T/prt
41 [the [horse T/prt]]
[the [horse T/prt]] + race
42 [the [horse T/prt(V)]]
[the [horse T/prt(V)]] + past
43 [the [horse [T/prt(V) past]]]
[the [horse [T/prt(V) past]]] + the
44 [the [horse [T/prt(V) [past the]]]]
[the [horse [T/prt(V) [past the]]]] + barn
45 [the [horse [T/prt(V) [past [the barn]]]]]
[the [horse [T/prt(V) [past [the barn]]]]] + T
46 [the [horse [T/prt(V) [past [[the barn] T]]]]]
[the [horse [T/prt(V) [past [[the barn] T]]]]] + fell
47 [the [horse [T/prt(V) [past [the [barn T]]]]]]
[the [horse [T/prt(V) [past [the [barn T]]]]]] + fell
48 [the [horse [T/prt(V) [[past [the barn]] T]]]]
[the [horse [T/prt(V) [[past [the barn]] T]]]] + fell
49 [the [horse [[T/prt [race [past [the barn]]] T]]]
[the [horse [[T/prt [race [past [the barn]]] T]]]] + fell
50 [the [horse [[T/prt [race [past [the barn]]] T fell]]]] (<= accepted)

```

The exact derivation is sensitive to the actual parsing principles that are activated before the simulation trial. For example, if we assume that the participle verb is activated before the finite

verb (against experimental data), then the model will reach the correct solution without garden paths.

We do not assume that the backtracking operation visible in the example above is entirely realistic from the psycholinguistic point of view. There are two reasons why it exists. One is that by performing systematic backtracking we let the model to explore the complete parsing tree. If we only generated the first solution, spurious secondary solutions might escape our attention. It is not untypical that the model finds ungrammatical and/or wrong secondary solutions, revealing that it is using a wrong competence model. The second reason is that backtracking gives us important information concerning the model's performance. We can compare the amount of computational resources spend in processing different types of constructions and/or different algorithmic solutions. In addition, realistic language comprehension by the human brain in connection with canonical sentences is seldom subject to any garden pathing, so we can at least aim for a model that finds the correct solution immediately, at least in those cases.³³

There is one situation in which incrementality is violated: inflectional features are stored in a temporary working memory buffer and enter the syntactic component inside lexical items. The third person agreement marker -s in English, for example, will be put into a temporary memory hold, inserted inside the next lexical item, which then enters syntax. The element therefore stays in the memory a very brief moment, being discharged as soon as possible. In the case of several inflectional features, they are all stored in the same memory system and then inserted inside the next lexical item as a set of features (order is ignored).

³³ It is possible, in fact a worthwhile goal, to add a realistic backtracking model on the top of the systematic backtracking as an optional component. I suspect that real speakers solve garden path problems by starting the parsing from the beginning with additional noise added to the decision mechanism.

5.3.2 *Connectness*

Connectness refers to the property that all incoming linguistic material is attached to one phrase structure that connects everything together. There never occurs a situation where the syntactic working memory holds two or more phrase structures that are not connected to each other by means of some grammatical dependency. Adjunct structures are attached to their host structures loosely: they are geometrical constituents inside their host constructions, observing connectness, but invisible to many computational processes applied to the host (§ 4.6). We can imagine of them being pulled out from the computational pipeline processing the host structure and being processed by an independent computational sequence. Each adjunct, and more generally phase, is transferred independently and enters the LF-interface as an independent object.

5.3.3 *Seriality*

Seriality refers to the property that all operations of the parser are executed in a well-defined linear sequence. Although this is literally true of the algorithm itself, the detailed serial algorithmic implementation cannot be mapped directly into a cognitive theory for several reasons. One reason is that each linguistic computational operation performed during processing is associated with a predicted cognitive cost measured in milliseconds, and it is the linear sum of these costs that provides the user the predicted cognitive processing time for each word and sentence. The current parsing model is serial in the sense that the predicted cognitive cost is computed in this way by adding the cognitive costs from each individual operation together. We could include parallel processing into the model by calculating the cognitive cost differently, i.e., not adding up the cost of computational operations that are predicted to be performed in parallel. Another complicating factor is that in some cases the implementation order does not seem to matter. We could simply *assume* that the processing is implemented by utilizing parallel processes. Despite these concerns, most computational operations implemented by the linear phase model must be executed in an exact specific order in order to derive empirically correct results. Therefore, for the most part the model assumes that language processing is serial.

5.3.4 Locality preference

Locality preference is a heuristic principle of the human language comprehension which requires that local attachment solutions are preferred over nonlocal ones. A local attachment solution means the lowest right edge in the existing partial phrase structure representation. Thus, the preposition phase *with a telescope* in (61) is first attached to the lowest possible solution, and only if that solution fails, to the nonlocal node.

- (61) John saw the girl with a telescope.

It is possible to run the parsing model with five different locality preference algorithms. They are as follows: *bottom-up*, in which the possible attachment nodes are ordered bottom-up; *top-down*, in which they are ordered in the opposite direction ('anti-locality preference'); *random*, in which the attachment order is completely random; *Z*, in which the order is bottom first, top second, then the rest in a bottom-up order; *sling*, which begins from the bottom node, then tries the top node and explores the rest in a top-down manner. The top-down and random algorithm constitute baseline controls that can be used to evaluate the efficiency of more realistic principles. The selected algorithm is defined for each independent study. If no choice is provided, bottom-up algorithm is used by default.

5.3.5 Lexical anticipation

Lexical anticipation refers to parsing decisions that are made on the basis of lexical features. The linear phase parser uses several lexical features (§ 4.3, 6.2.3). The system works by allowing each lexical feature to vote each attachment site either positively or negatively, and the sum of the votes will be used to order the attachment sites. The weights, which can be zero, can be determined as study parameters. This allows the researcher to determine the relative importance of various lexical features and, if required, knock them out completely (weight = 0). Large scale simulations have shown that lexical anticipation by both head-complement selection and head-specifier selection increases the efficiency of the algorithm considerably.

5.3.6 Left branch filter

The left branch filter closes parsing paths when the left branch constitutes an unrepairable fragment. The principle operates before other ranking principle are applied. The left branch filter can be turned on and off for each study.

5.3.7 Working memory

There is psycholinguistic evidence that the operation of the human language comprehension module is restricted by a working memory bottleneck. After considerable amount of simulation and exploration of other models I concluded that this hypothesis must be correct. The user can activate the working memory or knock it off by changing the study parameters and in this way see what its effects are.

Each constituent (node in the current phrase structure) is either active in the current working memory or inactive and out of the working memory. Any constituent β that arrives from the lexical component into syntax is active, and any complex constituent $[\alpha \beta]$ created thereby will be active. A constituent is inactivated and put out of the working memory when it is transferred and passes the LF-interface or when the attachment $[\alpha \beta]$ is rejected by ranking and/or by filtering; otherwise, it is kept in the working memory. It is assumed that a constituent residing out of the working memory is not processed in any way. All filtering and ranking principles cease to apply to it. Finally, it is assumed that re-activation of a dormant constituent accrues a cognitive cost, 500ms in this version. This implies that exploration of rejected parsing solutions will accrue much higher cognitive cost than they otherwise would do, as such dormant constituent must be reactivated into the working memory.

5.3.8 Conflict resolution and weighting

The abovementioned principles can conflict. It is possible, for example, that locality preference and lexical anticipation provide conflicting results. The conflict is solved by assuming that locality preference defines default behavior that is outperformed by lexical anticipation if the two are in conflict. When different lexical features provide conflicting results, it is assumed that they

cancel each out symmetrically. Thus, if head-complement selection feature votes against attachment $[\alpha \beta]$ but specifier selection favors it, then these votes cancel each other out, leaving the default locality preference algorithm (whichever algorithm is used). If two lexical features vote in favor, then the solution receives the same amount of votes as it would if only one positive feature would do the voting (+/- pair cancelling each other out, leaving one extra +). This result depends in how the various feature effects are weighted, which can again be provided independently for each study.

5.4 Measuring predicted cognitive cost of processing

Processing of words and whole sentences is associated with a predicted cognitive cost, measured in milliseconds. This is done by associating each word with a preprocessing time depending on its phonetic length (currently 25ms per phoneme) and then summing predictive cognitive costs from each computational operation together. Most operations are currently set to consume 5ms, but the user can define these in a way that best agrees with experimental and neurobiological data. Reactivation of a constituent that is not inside the active working memory consumes 500ms. The resulting timing information will be visible in the log files and in the resource outputs. The processing time consumed by each sentence is simply the sum of the processing time of all of its words. A useful metric in assessing the relative processing difficulty of any given sentence is to calculate the mean predicted cognitive processing time per each word (total time / number of words). This metric takes sentence length into account. Resource consumption is summarized in the resource output file (§ 6.3.5) that lists each sentence together with the number of all computational operations (e.g., Merge, Agree, Move) consumed from the reading of the first word to the outputting of the first legible solution. The file uses CSV format and can be read into an analysis program (Excel, SPSS, Matlab) or processed by using external Python libraries such as pandas or NumPy.

5.5 A note on implementation

Most of the performance properties are implemented in their own module *plausibility_metrics.py* which determines how the attachment solutions (§ 4.1) are filtered and ordered. The abstract linear parser, which does not implement any performance properties by itself, sends all available attachment solutions to this module, which will first focus the operation to those nodes which are in the active working memory; the rest are not processed. The active nodes are then filtered, so that only valid solutions remain. The user can knock off all filters by changing the parameters of the study. The remaining nodes are ordered by applying the selected locality preference algorithm, which provides the default ordering, and then by applying all other ranking principles such as lexical anticipation. Finally, the concatenated list of ordered nodes + nodes not active in the working memory and not processed are returned. The parser uses the ordered list to organize the parsing derivation. Notice that the inactive nodes that are not in the active working memory must be part of the list as the parser must be able to explore them if everything else fails, but since the plausibility module does not process them, their order is independent of the incoming word and the partial phrase structure representation currently being constructed.

6 Inputs and outputs

6.1 Installation and use

The program is installed by cloning the whole directory from Github.

```
https://github.com/pajubrat/parser-grammar
```

The user must define a folder in the local computer where the program is cloned. This folder will then become the root folder for the project. The root folder will contain at least the following subfolders: */docs* (documentation, such as this document), */language data working directory* (where each individual study is located) and */lpparse* (containing the actual Python modules). In order to run the program the user must have Python (3.x) installed on the local computer. Refer to the Python installation guide (www.python.org). The program can then be used by opening a command prompt into the program root folder and writing “python lpparse” into the command prompt as shown below:



```
Microsoft Windows [Version 10.0.19045.4291]
(c) Microsoft Corporation. Kaikki oikeudet pidätetään.

J:\Research>python lpparse
```

This will parse all the sentences from the designated test corpus file in a study folder. Each trial run that is launched by the above command involves a host of internal parameters that the user can configure. These involve things such as where is the lexicon, test corpus, what heuristic

principles should be used, and others. This information is defined in the following way. The root directory contains a file `$app_settings.txt`.

📁 .git	14.4.2024 12.34
📁 .idea	15.4.2024 9.11
📁 docs	10.6.2023 11.08
📁 language data working directory	4.4.2024 10.39
📁 lpparse	15.4.2024 9.39
📝 \$app_settings.txt	15.4.2024 9.28
📄 .gitignore	21.1.2024 11.24
📝 dev_log.txt	15.4.2024 9.39
📄 LICENSE	19.2.2022 10.33
📄 README.md	9.4.2024 20.42



This file contains a pointer to a *study configuration file* (*.lpg). For example:

```
study_folder=language data working directory/study-15-b-agreement  
study_file=agreement.lpg
```

In this case we use the study configuration file `agreement.lpg` (see Section 6.2.1) that is located in the *study folder* provided in the first line.

6.2 Structure of the input files

6.2.1 Study configuration file

6.2.1.1 General

Each study is defined by its own *study configuration file* that sets the parameters in a way intended by the researcher. The study configuration file is a list of parameters, each containing a key and a value separated by =. For example, the study configuration file `agreement.lpg` begins with the following key, value pairs:

```

# METADATA
meta_author= Anon
meta_year= 2024
meta_date= April
meta_study_id= 15
meta_comments=None

# FILES
file_study_folder=language data working directory/study-15-b-agreement
file_test_corpus=agreement_corpus.txt
file_lexicon_folder=language data working directory/lexicons
file_lexicons=lexicon.txt;UGmorphemes.txt;inflections.txt
file_redundancy_rules=redundancy_rules.txt

```

The key determines the name of the parameter. For example, the key *file_study_folder* determines the name of the parameter that defines the folder from where the program tries to find the test corpus file. It is followed by the name of the folder. The parameters are the following.

6.2.1.2 *Metadata*

Metadata consists of fields that define things like authorship, date and time, the numerical identity of the study and comments. The user can add any metadata here, provided that the parameter begins with “meta_”. This information is printed to the beginning of the results file.

```

# METADATA
meta_author= Anon
meta_year= 2024
meta_date= April
meta_study_id= 15
meta_comments=None

```

6.2.1.3 *Files*

The files section defines the names and locations of the external input files used in the simulation.

```

# FILES
file_study_folder=language data working directory/study-15-b-agreement
file_test_corpus=agreement_corpus.txt
file_lexicon_folder=language data working directory/lexicons
file_lexicons=lexicon.txt;UGmorphemes.txt;inflections.txt
file_redundancy_rules=redundancy_rules.txt

```

There can be any number of lexicon files.

6.2.1.4 General simulation parameters

```
# GENERAL SIMULATION PARAMETERS
general_parameter_only_first_solution = False
general_parameter_logging = True
general_parameter_use_numeration = False
general_parameter_calculate_assignments = False
general_parameter_calculate_pragmatics = True
general_parameter_calculate_thematic_roles = True
general_parameter_project_objects = True
general_parameter_calculate_predicates = True
```

Simulation parameters can be either True or False. They have the following interpretations:

only_first_solution will stop the derivation when the first legitimate output analysis has been found. This speeds up processing when examining observational adequacy. It will not find ambiguities, however. *logging* will set logging on or off. *use_numeration* will generate the input dataset from a numeration. *assignments* will attempt to calculate assignments and is involved in binding. *pragmatics* calculates information structure and other pragmatical attributes. *thematic roles* calculates thematic roles. *project_objects* will project semantic objects into the global discourse inventory (ontology). *predicates* will calculate predicate-argument dependencies.

6.2.1.5 Image parameters

Image parameters determine how phrase structure images are generated. Currently these images can only be generated by using the graphical user interface (Chapter 9).

```
# IMAGE PARAMETERS
image_parameter_words = True
image_parameter_glosses = True
image_parameter_adjuncts = False
image_parameter_complex_heads = False
image_parameter_head_chains = True
image_parameter_trivial_head_chains = False
image_parameter_phrasal_chains = True
image_parameter_DP_hypothesis = True
image_parameter_chain_curvature = 1
image_parameter_features = !∅LF,∅PF; +∅LF,∅PF; ∅LF; ∅PF; ARG; iPHI:PER:1; iPHI:PER:2; iPHI:PER:3
image_parameter_visualization = !∅LF,∅PF > ∅; +∅LF,∅PF > ∅; ?∅LF,∅PF > ∅; ∅PF > /∅/; ARG > ∅_
image_parameter_visualization = iPHI:* > i∅
```

6.2.1.6 UG components

This section determines certain parameters of the underlying linguistic theory of grammar. Currently the user can run the simulation either with the standard or revised theory of Agree. In addition, the user can also determine the phase heads here.

```
# UG COMPONENTS
# Agree
# standard = standard theory (Chomsky 2000, 2001, 2008)
# revised = current best formulation
UG_parameter_Agree = revised

# Phase theory
UG_parameter_phase_heads = C;v;FORCE;Inf;IMPSS
UG_parameter_phase_heads_exclude = φ;v-;T0/inf
```

6.2.1.7 Parsing heuristics

Parsing heuristics set a range of parameters that have to do with the heuristic principles the algorithm uses when attempting to find solutions for the input.

```
# PARSING HEURISTICS
extra_ranking = True
filter = True
lexical_anticipation = True
working_memory = True
closure = Bottom-up
positive_spec_selection = 100
negative_spec_selection = -100
break_head_comp_relations = -100
negative_tail_test = -100
positive_head_comp_selection = 100
negative_head_comp_selection = -100
negative_semantics_match = -100
lf_legibility_condition = -100
negative_adverbial_test = -100
positive_adverbial_test = 100
```

6.2.2 Test corpus file (any name)

The test corpus file name and location are provided in the *study configuration file*. Certain conventions must be followed when preparing the file. Each sentence is a linear list of phonological words, separated by white space and following by next line (return, or \n), which ends the sentence. Words that appear in the input sentences must be found from the lexicon exactly in the form they are provided in the input file. For example, do not use several forms (e.g.

admire vs. *Admire*) for the same word, do not end the sentence with punctuation, and so on.

Depending on the research agenda a fully disambiguated lexicon could be considered an option.

Special symbols are used to render to output more readable and to help testing. Symbols # and single quotation (‘) in the beginning of the line are read as introducing comments and are ignored. This allows the user to write glosses below the test sentences. Symbol & is also read as a comment, but it will appear in the results file as well. This allows the user to leave comments into the results file that would otherwise get populated with raw data only. Below is a screenshot from one test corpus file.

```
pesi#[foc] Pekka ikkunat
'wash.pst.3sg.foc Pekka.nom window.pl.acc'

pesi#[k0] Pekka ikkunat
'wash.pst.3sg.q Pekka.nom window.pl.acc'

*Pekka pesi#[k0] ikkunat
'Pekka.nom wash.pst.3sg.q window.pl.acc

Pekka antoi kirjan Merjalle
'Pekka.nom give.pst.3sg book.acc Merja.all'
```

Notice that one of the input sentences is marked as ungrammatical by prefacing the sentence with an asterisk*. A sentence that has not been marked by * will be assumed to be grammatical. These markings should correspond to the gold standard/native judgments. The parsing algorithm will ignore this symbol, but later it will be compared with the grammaticality judgment predicted by the model. Mismatches are listed into their own error file and printed into the console after all sentences have been processed. This allows the researcher to see immediately how successful the model is.

If a line is prefaced with %, the main script will process only that sentence. This functionality is useful if the user wants to examine the processing of only one sentence (input sentences can also be provided from the command prompt).

```

pesi#[foc] Pekka ikkunat
'wash.pst.3sg.foc Pekka.nom window.pl.acc'

% pesi#[k0] Pekka ikkunat
'wash.pst.3sg.q Pekka.nom window.pl.acc'

*Pekka pesi#[k0] ikkunat
'Pekka.nom wash.pst.3sg.q window.pl.acc

Pekka antoi kirjan Merjalle
'Pekka.nom give.pst.3sg book.acc Merja.all'

```

If the user wants to examine a group of sentences, they should all be prefaced with + symbol. Command =STOP= at the beginning of a line will cause the processing to stop at that point, allowing the user to process only *n* first sentences. To begin processing in the middle of the file, use the symbol =START= (in effect, sentences between =START= and =STOP= will be processed).

It is possible to feed the input sentences to the model as individual sentences or as part of a larger conversation. A conversation is defined as a sequence of sentences which share the same global discourse inventory. To create a conversation between two sentences, use semicolon at the end of the first sentence. The result of this is that the semantic objects instructed by the first sentence will be available as denotations for the expressions in the second (62).

- (62) a. John₁ met Mary₂;
- b. He_{1,3} admires her_{2,4}.

Any number of sentences can be sequences into a conversation. If the sentence does not end with a semicolon, it is assumed that the conversation ended and the global discourse inventory is reset when the next sentence is processed. Thus, if sentence (62)a did not end with the semicolon, pronouns *he* and *her* in sentence (b) can no longer refer to them. Conversations can be used to create discourse contexts for test sentences.

6.2.3 Lexical files

The main script uses three lexical resource files that are by default called *lexicon.txt*, *redundancy_rules.txt*, *ug_morphemes.txt* and *inflections.txt*. The first contains language specific lexical items, the second a list of universal redundancy rules, universal morphemes and inflectional affixes. Figure 15 illustrates the language-specific lexicon.

```
177 lähtemässä* :: lähte-#MA.ine/inf*#P* LANG:FI
178 lähtemästä :: lähte-#MA.elä/inf LANG:FI
179 lähtemättä :: lähte-#MA.abe/inf LANG:FI
180 lähtemällä :: lähte-#MA.adé/inf LANG:FI
181 lähteäkse :: lähte-#KSE/inf LANG:FI
182 lähteäkseen :: lähte-#KSE/inf#An LANG:FI
183 lähteäkseen* :: lähte-#A/inf*#KSE/inf#An LANG:FI
184 lähtien :: lähte-#E/inf LANG:FI
185 lähtiessä :: lähte-#ESSA/inf LANG:FI
186 lähtiessään :: lähte-#ESSA/inf#An LANG:FI
187 lähdettyääñ :: lähte-#TUA/inf#An LANG:FI
188 lähdettyä :: lähte-#TUA/inf LANG:FI
189 lähde :: lähte-#T/fin LANG:FI
190 lähde :: lähte-#T/default LANG:FI
191 lähti :: lähte-#T/fin#[V] LANG:FI
192 lähdettiin :: lähte-#impass#T/fin#tVVn LANG:FI
193 lähte- :: PF:lähte- LF:leave CLASS:INTR V SPEC:∅ -COMP:∅ COMP:P LANG:FI
194 lle :: PF:P(-lle) LF:for P -E +SEM:directional -∅PF TAIL:V LANG:FI
195 lla :: PF:P(-lla) LF:by P -∅PF TAIL:T TAIL:*V LANG:FI
196 lukemalla :: luke-#MA.adé/inf LANG:FI
197 lukee :: luke-#v#T/fin#[V] LANG:FI
```

Figure 15. Structure of the lexical file (*lexicon.txt*)

Each line in the lexicon file begins with the surface entry that is matched in the input. This is followed by :: which separates the surface entry from the definition of the lexical item itself. If the surface entry has morphological decomposition, it follows the surface entry and is given in the format ‘*m#m#m#...#m*’ where each item *m* must be found from the lexicon. Symbol # represents morpheme boundary. The individual constituents are thus separated by symbol # which defines the morphological decomposition. If the element designates a primitive (terminal) lexical item, the entry is followed by a list of lexical features. Each lexical feature will be inserted as such inside that lexical item, in the set constituting that item, when it is streamed into syntax. An inflectional feature is designated by the fact that its morphemic decomposition is replaced with symbol “–” or by the word “inflectional.” They are otherwise defined as any other lexical item, namely as a set of features. These features are inserted inside full lexical items during lexical streaming. The following screenshot shows the definitions for Finnish finite agreement affixes.

```

n      :: inflectional ØPF weak_pro PHI:NUM:SG,PER:1 PER PHI:DET:DEF LANG:FI
t      :: inflectional ØPF weak_pro PHI:NUM:SG,PER:2 PER PHI:DET:DEF LANG:FI
[V]    :: inflectional ØPF weak_pro nonreferential_pro PHI:NUM:SG,PER:3 PER LANG:FI
mme   :: inflectional ØPF weak_pro PHI:NUM:PL,PER:1 PER PHI:DET:DEF LANG:FI
tte   :: inflectional ØPF weak_pro PHI:NUM:PL,PER:2 PER PHI:DET:DEF LANG:FI
vAt   :: inflectional ØPF weak_pro PHI:NUM:PL,PER:3 PER LANG:FI
Vn    :: inflectional ØPF weak_pro PHI:NUM:PL,PER:1 HUM:HUM PHI:DET:DEF PER IMPSS LANG:FI

```

A lexical feature is a string, essentially a formal pattern. The way any given feature reacts inside syntax and semantics is defined by the computations that process these lexical items and the patterns in them.³⁴ The file *ug_morphemes.txt* is structured in the same way but contains universal morphemes such as T and v. The following screenshot shows the definitions for some of Finnish derivational morphemes.

```

# Inner layer (layer 1)

# Verbal

TA      :: V CAU ε          PF:-TA    LF:cause LANG:FI V/TR
ile,ele,skele :: V FREQ ε  PF:-ile   LF:frequent LANG:FI -wCOMP:FREQ,EVE2,MOM2
U,UTU   :: V REFL ε        PF:-utu   LF:become LANG:FI -wCOMP:FREQ,V/INTR,EVE2 V/INTR
AhtA   :: V MOM1 ε        PF:-AhtA  LF:moderated LANG:FI !wCOMP:V,V/INTR -wCOMP:ε
Aise   :: V MOM2 ε        PF:-Aise   LF:moderated LANG:FI !wCOMP:V,V/TR -wCOMP:ε
a      :: a-#v             PF:-a     LANG:FI
a-     :: V EVNT ε        PF:-a     LF:event LANG:FI !wCOMP:N -wCOMP:ε,n,a

# Nominal

O,nt0   :: N EVE ε        PF:-O     LF:action  LANG:FI !wCOMP:V -wCOMP:CAU,REFL,MOM1
Us,Ude :: N EVE2 ε       PF:-Us    LF:action  LANG:FI !wCOMP:V -wCOMP:FREQ,MOM2
st0    :: N COL ε         PF:-st0   LF:group   LANG:FI !wCOMP:N -wCOMP:COL
(t)e   :: N RSLT ε       PF:-(t)e  LF:result  LANG:FI !wCOMP:V
lAinen,lAise :: N PER ε  PF:-lAinen LF:person  LANG:FI !wCOMP:N
lise   :: N ε              PF:-lise-  LF:owner   LANG:FI !wCOMP:N

# Adjectival

ise     :: A QLT ε        PF:-ise   LF:quality LANG:FI -wCOMP:V,EVE2
mAise  :: A QLT ε        PF:-mAise  LF:quality LANG:FI !wCOMP:N

```

6.2.4 Lexical redundancy rules

Lexical redundancy rules are provided in the file *redundancy_rules.txt* and define default properties of lexical items unless otherwise specified in the language-specific lexicon. Redundancy rules are provided in the form of an implication ‘ $\{f_0, \dots, f_n\} \rightarrow \{g_1, \dots, g_n\}$ ’ in which

³⁴ They are currently implemented by simple string operations, but in some later iteration all such processing will be replaced by regex processing.

the presence of a triggering or antecedent features $\{f_0, \dots, f_n\}$ in a lexical item will populate features g_1, \dots, g_n inside the same lexical item. Below is a screenshot from the file containing the lexical redundancy rules.

```

5  FORCE   :: -ARG -EF Fin !COMP:!* !PROBE:Fin COMP:T/fin COI
6  C/fin    :: -ARG Fin C -COMP:C/fin -COMP:T/prt COMP:T/fin
7  C        :: -SPEC:MA/A EF
8  T/fin    :: ARG EF !SELF:∅ T Fin !COMP:!* -SPEC:C SPEC:T/p:
9  T/prt    :: ARG Fin T EF !COMP:!* COMP:v COMP:V COMP:T/prt
10 T       :: ARG COMP:!* COMP:v COMP:V -COMP:∅ -COMP:D -COMP:
11 Neg/fin  :: ARG EF Fin NEG COMP:T/prt COMP:T -SPEC:T -SPEC:
12 v       :: ARG -∅PF ASP !COMP:!* COMP:V -SPEC:v -SPEC:T/f:
13 D       :: ∅ -ARG !COMP:!* -COMP:D -COMP:v -COMP:V -COMP:
14 D/rel    :: ∅ D -ARG OP OP:REL COMP:!* -COMP:C/fin -COMP:∅
15 N       :: -COMP: -COMP:EXPL -COMP:VA/inf -COMP:T/fin -I
16 V       :: ARG -∅PF ASP -SPEC:T/fin -SPEC:FORCE SPEC:Adv
17 P       :: ARG COMP:∅ -COMP:V -COMP:MA/A -SPEC:Neg -COMP:
18 Adv     :: -SPEC:N -SPEC:FORCE -SPEC:Neg/fin -SPEC:T/fin
19 A       :: COMP:∅ TAIL:∅ -SPEC:A
20 Inf     :: TAM -COMP:T/fin -COMP:FORCE -COMP:C/fin -SPEC
21 n       :: ∅ EF ARG Inf !COMP:!* COMP:∅ COMP:V COMP:v
22 ∅       :: SPEC:∅ 2SPEC -COMP:∅
23 Num     :: ∅ COMP:∅ COMP:N COMP:n -COMP:D -COMP:QN
24 DET     :: ∅ COMP:∅ -COMP:DET
25 QN      :: ∅ COMP:∅ COMP:N COMP:Num
26 ∅       :: COMP:N -SPEC:T -SPEC:P COMP:D/rel !COMP:*
.

```

The antecedent features are written to the left side of the :: symbol, and the result features to the right. Both feature lists are provided by separating each feature (string) by whitespace. The lexical resources are processed so that the language-specific sets are created first, followed by the application of the lexical redundancy rules. If a lexical redundancy rule conflicts with a language-specific lexical feature, the latter will override the former. Thus, lexical redundancy rules define the “default features” associated with any given triggering feature. It is also possible (and in some cases needed) to use language specific redundancy rules. These are represented by pairing the antecedent feature with a language feature (e.g., [LANG:FI]).

6.3 Structure of the output files

6.3.1 Errors

A list of sentences where the model erred will be in the *error reports* file. If this file is empty, then the model is observationally adequate.

6.3.2 Results

The name and location of the results file is determined when configuring the main script by the study configuration file. The default name is made up by combining the test corpus name together with `_results.txt`. Each time the main script is run, the default results file is overridden. The file begins with list of study parameters used, time of the execution and locations of the input files. This followed by a grammatical analysis and other information concerning each example in the test corpus, each provided with a numeral identifier. What type of information is visible depends on the aims of the study.

6.3.3 The derivational log file

The derivational log file, created by default by adding `_log` into the name of the test corpus file, contains a more detailed report of the computational steps consumed in processing each sentence in the test corpus and of the semantic interpretation. The log file uses the same numerical identifiers as the results file. By default the log file contains information about the processing and morphological decomposition of the phonological words in the input, application of the ranking principles leading into Merge-1, transfer operation applied to the final structure when no more input words are analyzed and many aspects of semantic interpretation. Intermediate left branch transfer operations are not reported in detail. Each section in the derivational log file begins with the input sentence as specified by the linear list of phonological words it contains:

```
Sentence #2014. John admires Mary

Phonological words: ['John', 'admires', 'Mary']
Sensory Processing of /#John/
Sensory Processing of /#admires/
Sensory Processing of /#Mary/
```

Sensory processing is cited here since it adds to the total resource consumption but currently does nothing else. The first word `/John/` is then processed as follows. The word retrieves a morphological chunk that is decomposed into several morphemes, the first which are inflectional and are mapped into feature sets.

```

Next morpheme /John/ retrieves (1) morphological chunk [John-#D#sg#3p#def#hum#m]
  Next morpheme /#m_-/ retrieves (1) m = [LANG:EN][iPHI:GEN:M][inflectional]
  Next morpheme /#hum#/ retrieves (1) hum = [LANG:EN][iPHI:HUM:HUM][inflectional]
  Next morpheme /#def#/ retrieves (1) def = [LANG:EN][iPHI:DET:DEF][inflectional]
  Next morpheme /#3p#/ retrieves (1) 3p = [LANG:EN][iPHI:PER:3][inflectional]
  Next morpheme /#sg#/ retrieves (1) sg = [LANG:EN][iPHI:NUM:SG][inflectional]
  Next morpheme /#D#/ retrieves (1) D

```

Finally D and N are encountered that retrieve two separate zero-level heads from the lexicon:

Next head D⁰

```
Next morpheme /_John-#/ retrieves (1) John-
```

Next head John-⁰

```

Ranking:
(1) [D↓+ N°]
Merge((N D)0) => (N D)0 $0x1$
```

Notice that D⁰ does not cause any further syntactic operations since it is the first syntactic element in the syntactic working memory. N⁰ will be merged with D⁰ to create complex had (N D)⁰. The parsing then continues by applying the same process to each phonological word in the input string.

A small segment of this process:

Next head T/fin⁰

```

Ranking:
(1) [(N D)0↓+ T0] $0x2$
(2) [(N D)0 + T0] $0x3$
Merge([[D John] T]) => [[D John] T] $0x2$
```

```
Next morpheme /#v#/ retrieves (1) v
```

Next head v⁰

```

Ranking:
(1) [T↓+ v0] $0x4$
Merge([[D John] (v T)0]) => [[D John] (v T)0] $0x4$
```

```
Next morpheme /_admire-#/ retrieves (1) admire-
```

Next head admire-⁰

```

Ranking:
(1) [(v T)0↓+ V0] $0x5$
Merge([[D John] ((V v) T)0]) => [[D John] ((V v) T)0] $0x5$
```

```

Next morpheme /Mary/ retrieves (1) morphological chunk [Mary-#D#sg#3p#def#hum#proper_name#f]
Next morpheme /#f_-/ retrieves (1) f = [LANG:EN][iPHI:GEN:F][inflectional]
Next morpheme /#proper_name#/ retrieves (1) proper_name = [LANG:EN][R:NEW:$$/SUR][R:REF:NAME][i
Next morpheme /#hum#/ retrieves (1) hum = [LANG:EN][iPHI:HUM:HUM][inflectional]
Next morpheme /#def#/ retrieves (1) def = [LANG:EN][iPHI:DET:DEF][inflectional]
Next morpheme /#3p#/ retrieves (1) 3p = [LANG:EN][iPHI:PER:3][inflectional]
Next morpheme /#sg#/ retrieves (1) sg = [LANG:EN][iPHI:NUM:SG][inflectional]
Next morpheme /#D#/ retrieves (1) D
```

Once all words have been decomposed and their constituents merged-1 into the phrase structure in the syntactic working memory, the result will be transferred to the LF-interface (or LF-interface is reconstructed from the PF-interface):

```

PF-interface [[D John] (((V v) T)° (N D)°)]

Head Chain(D) => [[D John] ((V v) T)° [D Mary]]
Head Chain(T) => [[D John] [T [v [admire [D Mary]]]]]
Feature Inheritance(T) => [[D John] [T [v [admire [D Mary]]]]]
Phrasal Chain([D John]) => [[D John]:11 [T [__:11 [v [admire [D Mary]]]]]]
Agree(admire°, D) valued features [NUM:SG] [PER:3] [DET:DEF] [GEN:F]
Agree(v°, D) valued features [NUM:SG] [PER:3] [DET:DEF] [GEN:F]
Agree(T°, D) valued features [NUM:SG] [PER:3] [DET:DEF] [GEN:M]

LF-interface [[D John]:11 [T [__:11 [v [admire [D Mary]]]]]]

```

This shows head chains, A-chain, \bar{A} -chains and agreement operations, in the order they were performed during transfer. Various LF-interface legibility tests and post-syntactic interpretation tests are then applied to the LF-interface. If they pass, the output is accepted:

```

LF-interface and postsyntactic legibility tests...
Argument for T°: pro, indexed to [D John]
Project [D John]: (1, QND)(2, GLOBAL)
Argument for v°: pro, indexed to [D Mary]
Argument for admire°: pro, indexed to [D Mary]
Project [D Mary]: (3, QND)(4, GLOBAL)
Calculating information structure...
Accepted.

```

This is then followed by a list of lexical items at the LF-interface and their feature content

Lexical features:	
D	[D][∅][LF:D][PF:D][−ARG][iPHI:DET:DEF][iPHI:GEN:M][iPHI:GEN:F]
John	[N][LF:John][PF:John][−COMP:A][−COMP:AUX][−COMP:D][−COMP:F]
T	[Fin][T][LF:T][PF:T][EF][ARG][Fin][PHI:DET:DEF][PHI:GEN:M][iPHI:GEN:F]
D	[D][∅][LF:D][PF:D][−ARG][iPHI:DET:DEF][iPHI:GEN:M][iPHI:GEN:F]
John	[N][LF:John][PF:John][−COMP:A][−COMP:AUX][−COMP:D][−COMP:F]
v	[V][v][LF:v][PF:v][ARG][ASP][PHI:DET:DEF][PHI:GEN:F][iPHI:GEN:F]
admire	[V][LF:admire][PF:admire][ARG][ASP][PHI:DET:DEF][PHI:GEN:M][iPHI:GEN:F]
D	[D][∅][LF:D][PF:D][−ARG][iPHI:DET:DEF][iPHI:GEN:F][iPHI:GEN:M]
Mary	[N][LF:Mary][PF:Mary][−COMP:A][−COMP:AUX][−COMP:D][−COMP:F]

followed by the same information that is available in the results file (ignored here), followed by a list of objects that were projected into the global discourse inventory during the processing:

Ontology:

```
Object 2 in GLOBAL: [D John]
Object 4 in GLOBAL: [D Mary]
```

which is followed by some their more detailed properties. Then the algorithm will attempt to find alternative solution by backtracking (using the rankings established during first-pass parsing):

Checking if the sentence is ambiguous...

```
Backtracking...
Backtracking...
```

6.3.4 Simple logging

A file ending with *_simple_log.txt* contains a simplified log file which shows only a list of the partial phrase structure representations and accepted solutions generated during the derivation.

6.3.5 Resources file

The algorithm records the number of computational operations and CPU resources (in milliseconds) consumed during the processing of the first solution. At the present writing, these data are available only for the first solution discovered. Recursive and exhaustive backtracking after the first solution has been found, corresponding to a real-world situation in which the hearer is trying to find alternative parses for a sentence that has been interpreted, is not relevant or psycholinguistically realistic to merit detailed resource reporting. These processed are included in the model only to verify that the parser is operating with the correct notion of competence and does not find spurious solutions. In addition, resource consumption is not reported for ungrammatical sentences, as they always involve exhaustive search.

6.3.6 Phrase structure images

It is possible to examine the output of the algorithm in the form of phrase structure images. This functionality is currently available through the graphical interface, see Chapter 9.

6.4 Inputs, outputs and the graphical user interface

It is possible to examine and analyze the inputs and outputs by using a graphical interface. This functionality is explained in Chapter 9.

DRAFT

7 Grammar formalization

7.1 Basic grammatical notions (*phrase_structure.py*)

7.1.1 Introduction

The phrase structure class defined in *phrase_structure.py* defines the phrase structure objects called constituents that are manipulated at each stage of the processing pipeline.

7.1.2 Lexical items

Lexical items are sets of lexical features. Once they enter syntax, they are associated with a phrase structure node and become *constituents*. We can think of the lexical component as providing feature sets wrapped inside constituents in the syntactic component. Phrase structure nodes define the phrase structure geometry.

7.1.3 Phrase structure geometry

The *daughter constituents* of any constituent α are provided in a list [X Y], which then becomes the *mother* of both daughters. It is assumed that there can be only two daughter constituents, the *left* and the *right*, which are defined as follows:

```
def left(X):
    if X.const:
        return X.const[0]

def right(X):
    if X.const:
        return X.const[-1]
```

The left constituent is therefore always the first constituent in the list, the right the last. From this it follows that if α has only one constituent, it will be both left and right. These constituents are called *affixes*:

```

def affix(X):
    if X.left() == X.right():
        return X.left()

```

If α has three constituents (possible in theory), then all other constituents except the first and the last will be invisible. We can then provide the following further definitions:

```

def terminal(X):
    return not X.const

def complex(X):
    return len(X.const) > 1

def zero_level(X):
    return len(X.const) < 2

```

A *terminal* constituent has no constituents, complex constituent has more than one, while a *zero_level* constituent has less than two. A terminal constituent is also a zero-level constituent.

7.1.4 Sisters

Two constituents are *geometrical sisters* if they have the same mother. The right constituent constitutes the geometrical sister of the left constituent, and vice versa.

```

def geometrical_sister(X):
    if X.is_left():
        return X.mother().right()
    return X.mother().left()

```

The notion of geometrical sister is defined in terms of phrase structure geometry. We will often use a narrower notion, called *sister*, that ignores externalized right adjuncts.

```

def sister(X):
    while X.mother():
        if not X.geometrical_sister().adjunct:
            return X.geometrical_sister()
        X = X.mother()

```

7.1.5 Proper selected complement

A (regular) *complement* is defined as the same relation as sisterhood (§ 7.1.4). A *proper selected complement* of a primitive constituent is its right sister.

```

def right_sister(X):
    if X.sister() and X.sister().is_right():
        return X.sister()

def proper_selected_complement(X):
    if X.zero_level():
        return X.right_sister()

```

7.1.6 Heads (*labels*), maximal projection, container

The recursive labeling algorithm § 4.5 is provided informally in (63).

(63) Labeling

Suppose α is a complex phrase. Then

- a. if the left constituent of α is primitive, it will be the label; otherwise,
- b. if the right constituent of α is primitive, it will be the label; otherwise,
- c. if the right constituent is not an adjunct, apply (14) recursively to it; otherwise,
- d. apply (14) to the left constituent (whether adjunct or not).

In Python:

```

def head(X):
    for x in [X] + X.const:
        if x.zero_level():
            return x
    return X.right_nonadjunct().head()

```

The notion of head of α allows us to define several other notions, such as *maximum projection*

```

def max(X):
    x = X
    while x.mother() and x.mother().head() == X.head():
        x = x.mother()
    return x

```

and the notions of *container* and *inside*, where container of α is the head of the mother of α and $\text{inside}(\alpha, \beta)$ determines whether α is a projection from β .

```

def inside(X, head):
    return X.head() == head

def container(X):
    if X.mother() and X.head() != X.mother().head():
        return X.mother().head()

```

Maximum projection therefore refers to the highest node dominating α that has the same head as α , and container refers to the head of α 's mother.

7.1.7 Minimal search, geometrical minimal search and upstream search

Several operations require that the phrase structure is explored in a pre-determined order. The *minimal search* from α explores the right edge of α in a downstream direction and crawls downwards on the right edge of α . It is defined by creating an iteration over phrase structure. The iterator is defined by the Python `__next__` and `__iter__` functions called when constructing iterators.

```

def __iter__(X):
    X.nn = X
    return X

def __next__(X):
    if not X.nn:
        raise StopIteration
    current = X.nn
    if X.nn.zero_level():
        X.nn = None
        return current
    elif X.nn.head() == X.nn.right().head() or X.nn.left().proper_selected_complement():
        X.nn = X.nn.right()
    else:
        X.nn = X.nn.left()
    return current.left()

```

Suppose we examine constituent α . Constructing an iterator over α calls `__iter__` which sets the next node = α and returns it. α will be the first item in the sequence. The iterator will then call `__next__`, which returns α if it is primitive and sets the next node = `None`, which terminates the iteration (first condition inside `__next__`). Suppose $\alpha = [A B]$. The next node in the iterator will be A (`return current.left()`). We must also decide whether the iterator goes inside A or B . It branches into B if and only if either (i) $[_{BP} A B]$ or (ii) $[A^0 B]$ (A^0 = primitive and B is a selected

right sister). If neither is true, it branches into A. We can now implement a “bare minimal search” by defining an iteration over α . The minimal search function in the current implementation is slightly more abstract and defined as follows:

```
def minimal_search(X, selection_condition=lambda x: True, sustain_condition=lambda x: True):
    """Collect items that satisfy selection_condition as long as sustain_condition remains true"""
    return takewhile(sustain_condition, (x for x in X if selection_condition(x)))
```

It has two additional properties: selection function and sustain condition. The former selects items from the bare minimal search based on the condition given in the input (*if selection_condition(const)*), while the latter can be used to define an intervention or termination that halts the search when the condition becomes true (*sustain_condition*). Both are provided as lambda functions. If no functions are provided, all constituents are returned and the search terminates only when a primitive constituent is reached as defined in the *__next__* function. *Geometrical minimal search* depends on the phrase structure geometry alone and does not respect labelling and visibility.

```
def geometrical_minimal_search(x):
    search_list = [x]
    while x.complex() and x.right():
        search_list.append(x.right())
        x = x.right()
    return search_list
```

7.1.8 Upward paths

The model makes much use of upward paths. Suppose that α constitutes a probe that triggers the scanning operation. First we define the notion of *upward path*.

```
def upward_path(x):
    upward_path = []
    x = (X, x.mother())
    while x[1]:
        if x[1].left() and x[1].left() != x[0]:
            upward_path.append(x[1].left())
        x = (x[1], x[1].mother())
    return upward_path
```

We proceed upwards and collect all left constituents into a list, ignoring constituents which dominate the upward path itself. *Edge* contains all elements in the path that are inside the projection from α .

```
def edge(x):
    return list(takewhile(lambda x: x.mother() and x.mother().inside(x), x.upward_path()))
```

7.1.9 Cyclic Merge

Simple Merge takes two constituents X, Y and yields [X, Y], X being the left constituent, T the right constituent. It is implemented by the class constructor `__init__()`, which takes X and Y as arguments and returns a new constituent. It sets up the mother-of relations for both sisters.

```
def __init__(self, left=None, right=None):
    self.const = []
    if left and right:
        self.create_constituents([left, right])

def create_constituents(X, lst):
    X.const = lst
    for x in X.const:
        x.mother_ = X
```

7.1.10 Countercyclic Merge-1

Countercyclic Merge-1 (`merge_1(α , β , direction)`) targets constituent α inside an existing partial phrase structure and creates a new constituent γ by merging β either to the left or right of α : $\gamma = [\alpha, \beta]$ or $[\beta, \alpha]$. Thus, if we have a phrase structure $[X \dots \alpha \dots Y]$, then Merge-1 generates either (a) or (b).

(64)

- a. $[X \dots [{}_\gamma \alpha \beta] \dots Y]$
- b. $[X \dots [{}_\gamma \beta \alpha] \dots Y]$

Constituent γ then replaces α in the phrase structure, with the phrase structural relations updated accordingly. Both Merge to the right and left, and both countercyclically and by extending the

structure, are allowed. The range of options is compensated by the restricted conditions under which each operation can occur. The fact that Merge-1 dissolves into separate processes is reflected in the code, which contains three separate functions: the first (*local_structure*) obtains a snapshot of the local structure around α (its mother and position in the left-right axis), the second creates $[\alpha \beta]$ or $[\beta \alpha]$ (*asymmetric_merge*), and the third (*substitute*) substitutes α with the new constituent $[\alpha \beta]$ by using local constituent relations recorded by the first operation.

```

def Merge_inside(X, C, direction=''):
    local_structure = (X.mother(), X.is_left())
    X = X.asymmetric_merge(C, direction)
    X.substitute(local_structure)
    return X

# Asymmetric Merge is a generalization of the bottom-up Merge (__init__) that can be
def asymmetric_merge(X, B, direction='right'):
    X.consume_resources('Merge-1', X)
    if direction == 'left':
        return PhraseStructure(B, X)
    return PhraseStructure(X, B)

def substitute(X, local_structure):
    if local_structure[0]:
        if not local_structure[1]:
            local_structure[0].create_constituents([local_structure[0].left(), X])
        else:
            local_structure[0].create_constituents([X, local_structure[0].right()])
    X.mother_ = local_structure[0]

```

7.1.11 Remove

An inverse of countercyclic Merge-1 is remove ($\alpha.remove$), which removes constituent α from the phrase structure and repairs the hole. This operation is used when the parser attempts to reconstruct movement: it merges elements into candidate positions and removes them if they do not satisfy a given set of criteria.

```

def remove(X):
    if X.mother():
        mother_, sister, grandmother = X.context()
        if mother_.is_right():
            grandmother.const = [grandmother.left(), sister]
        elif mother_.is_left():
            grandmother.const = [sister, grandmother.right()]
        X.mother_ = None

```

7.1.12 Feature checking

There are currently two feature checking operations, one which checks all features listed in the input set and another which checks some features listed in the input set.

```
def check(X, feature_set):
    return feature_set == {'*'} or feature_set <= X.head().features

def check_some(X, feature_set):
    return feature_set == {'*'} or feature_set & X.head().features
```

7.1.13 Probe-goal: *probe(label, goal_feature)*

Probe-goal relations are interpreted as downward pointing long-distance dependencies that are created by following minimal search. The function is currently used only to implement nonlocal selection. Suppose P is the probe head, G is the goal feature, and α is its (non-adjunct) sister in configuration [P, α], then we have define a probe-goal relation (65).

(65) Probe-goal

Under [P, α], G the goal feature, search for G from left constituents by going downwards inside α along its right edge by using minimal search.

For everything else than criterial features, G must be found from a primitive head to the left of the right edge node. The present implementation has an intervention clause which blocks further search if search encounters an intervention feature or a special condition is satisfied that involves probing for a tail features. The intervention system is not fully developed, and there is no systematic study concerning its properties.

```
def probe_goal_test(X):
    if X.sister():
        for ff in [(f, X.probe(f[7:])) for f in X.features if 'PROBE:' in f]:
            if (mandatory(ff[0]) and not ff[1]) or (illicit(ff[0]) and ff[1]):
                return True

def probe(X, G):
    return next((x for x in X.sister() if x.check({G})), None)
```

The probe-goal mechanism implements nonlocal selection. One example of nonlocal selection is the relationship between D and N. Every D must be paired with an N, but the relationship can be intervened by a number of other functional heads such as Q or Num.

7.1.14 Tail-head relations

A tail-head dependency is formed when a probe P must locate a goal G either inside the head of its own projection (“strong tail test”) or inside an upward path by memory scanning (21) (“weak tail test”). The tail test can be *positive*, in that it checks the existence of G, or *negative*, where it checks the absence of G (the test fails if G is present). Goals G are defined by *target features* $F = \{f_1 \dots f_n\}$ which must all be checked in order for the dependency to form. The main function is as follows.

```
def tail_test(X):
    pos_tsets = {frozenset(positive_features(tset)) for tset in X.get_tail_sets() if positive_features(tset)}
    neg_tsets = {frozenset(negative_features(tset)) for tset in X.get_tail_sets() if negative_features(tset)}
    checked_pos_tsets = {tset for tset in pos_tsets if X.tail_condition(tset)}
    checked_neg_tsets = {tset for tset in neg_tsets if X.tail_condition(tset)}
    return pos_tsets == checked_pos_tsets and not checked_neg_tsets
```

The function collects positive and negative target features and checks the tail condition for each. The test passes if all positive features are checked but none of the negative features are. The tail condition is defined as follows.

```
def tail_test(X):
    pos_tsets = {frozenset(positive_features(tset)) for tset in X.get_tail_sets() if positive_features(tset)}
    neg_tsets = {frozenset(negative_features(tset)) for tset in X.get_tail_sets() if negative_features(tset)}
    checked_pos_tsets = {tset for tset in pos_tsets if X.tail_condition(tset)}
    checked_neg_tsets = {tset for tset in neg_tsets if X.tail_condition(tset)}
    return pos_tsets == checked_pos_tsets and not checked_neg_tsets

def tail_condition(X, tset):
    if X.referential() or X.preposition():
        for m in (x for x in X.upward_path() if x.zero_level() and x.check_some(tset)):
            return m.check(tset)
    if not X.referential():
        return X.check_container(tset) or X.check_aunt(tset)

def check_container(X, tset):
    return X.max().container() and X.max().container().check(tset)

def check_aunt(X, tset):
    return X.max().mother().sister() and X.max().mother().sister().check(tset)
```

The test checks two different condition depending on whether the test its is referential (or preposition) or not. Nonreferential constituents must satisfy a stronger condition that is defined in the second clause and essentially tests that the tail features are present in the head of the projection inside of which the testing element is. This applies to adverbials, which are required to appear inside projections from specific heads. The second clause uses a weaker test and only requires that the tail feature can be found inside an upward path.

7.2 Transfer

7.2.1 A comment on the current version (15.0)

The organization and operation of the transfer function has undergone major refactoring and simplification, therefore this version should be regarded as the first version of a longer evolution. It will be documented here in its incomplete stage due to the fact that the current source code does not match at all with what was documented in previous versions.

Before examining the logic of the code, a few words towards the general philosophy behind the current approach. As discussed in § 4.8, transfer seems to be composed out of several operations, such as head reconstruction, A-chains, \bar{A} -chains and scrambling, that on the surface have different properties. For this reason in many earlier models each reconstruction operation was defined in its own class and utilized functions that were specific to that operation. It was simply impossible to calculate the data without using specialized modules for each operation. On the other hand, it was clear from the beginning that they were based on the same general template when examined from a more abstract point of view. Each operation was triggered by a deviance or error in the input and sought to correct it. An error in this context means a configuration or property in the spelled out structure (the structure that was submitted to transfer) that cannot pass LF-legibility and/or semantic interpretation. In the current version, some significant steps have been taken to unify all reconstruction operations under the same function, but the work is not complete; therefore, what exists at this stage is a two-factored system where all reconstruction operations utilize the same template but a residuum of differences remain that are defined elsewhere.

7.2.2 Generalized transfer

Reconstruction is a reflex-like operation that takes place without interruption from the beginning to the end. From an external point of view, it constitutes one step. The general form of each step is defined by *reconstruct* which takes two arguments, the starting node of each cycle, usually the bottom right node, and *instructions*. The *instructions* parameter is a dictionary which contains lambda-functions providing the conditions and operations that are observed and executed during

each reconstruction cycle, depending on the type of reconstruction. This dictionary, shown below, therefore contains most of the operation-specific information that remains in the current theory and which have not been yet reduced away.

```

instructions = {
    'Head': {'type': 'Head Chain',
              'test integrity': lambda x: x affix() and not x.right().copied and not x.License_EHM(),
              'repair': lambda x: x.create_chain(),
              'selection': lambda x: True,
              'sustain': lambda x: True,
              'legible': lambda x, y: y.properly_selected() and not y.empty_finite_EPP() and y.right_sister(),
              'single operation': False,
              'prepare': lambda x: x.prepare_head_chain(),
    'Phrasal': {'type': 'Phrasal Chain',
                'test integrity': lambda x: not x.copied and x.complex() and x.is_left() and not x.expletive(),
                'repair': lambda x: x.create_chain(),
                'selection': lambda x: x.zero_level() and not x.finite(),
                'sustain': lambda x: not (x.zero_level() and x.referential()),
                'legible': lambda x, y: x.Abar_legible(y),
                'prepare': lambda x: x.prepare_phrasal_chain(),
    'Feature': {'type': 'Feature Inheritance',
                'test integrity': lambda x: x.check({'ARG?'} or x.highest_finite_head()),
                'repair': lambda x: x.feature_inheritance(),
                'single operation': False,
    'A-chain': {'type': 'Phrasal Chain',
                'repair': lambda x: x.create_chain(),
                'selection': lambda x: x.has_vacant_phrasal_position(),
                'legible': lambda x, y: True,
                'sustain': lambda x: not (x.zero_level() and x.referential()),
                'test integrity': lambda x: not x.copied and x.complex() and x.is_left() and x.container() and not x.expletive(),
                'prepare': lambda x: x.prepare_phrasal_chain(),
    'Agree': {'type': 'Agree',
              'test integrity': lambda x: x.is_unvalued(),
              'repair': lambda x: PhraseStructure.speaker_model.Experimental_functions.Agree(x),
    'Extraposition': {'type': 'Extraposition',
                      'test integrity': lambda x: x.zero_level() and (x.top().contains_finiteness() or x.top().referential()),
                      'repair': lambda x: x.extrapose(),
    'Right Scramble': {'type': 'Right Scrambling',
                        'test integrity': lambda x: not x.copied and x.trigger_right_node_scrambling(),
                        'repair': lambda x: x.right_scrambling(),
    'Left Scramble': {'type': 'Left Scrambling',
                      'test integrity': lambda x: x.complex() and not x.copied and x.trigger_scrambling(),
                      'repair': lambda x: x.reconstruct_scrambling(),
    'Last Resort Extraposition': {'type': 'Last Resort Extraposition',
                                  'test integrity': lambda x: (x.top().contains_finiteness() or x.top().referential()),
                                  'repair': lambda x: x.last_resort_extrapose()}}
}

```

Crucially, the dictionary contains lambda functions which provide the operation-specific properties. In an ideal word they would not exist; rather, all reconstruction would use the same functions. The reconstruction cycle itself is defined as follows.

```

def reconstruct(X, op):
    for const in (x for x in [X.bottom()] + X.bottom().upward_path() if op['test integrity'](x)):
        op['repair'](const)
        X.speaker_model.results.record_derivation_step(const.top(), f'{op["type"]}: {const}')
        PhraseStructure.speaker_model.results.consume_resources(op['type'], const)

```

The function targets the starting node (usually the bottom node), examines if the node needs repair and, if it does, calls the repair function as specified in the instructions. Each *reconstruction cycle* moves from the bottom node towards the top node, moving from head to head, and performs the required repair operations for each node it finds to require repair. The nodes needing repair are detected by the *test integrity* entry in the instructions dictionary, a lambda function that is provided

by the caller and which again depends on the nature of the cycle. For example, head reconstruction is triggered whenever the node is a complex head.

Because transfer is executed in one well-ordered sequence, as shown above, this loop is executed several times, once for every reconstruction type. An alternative, perhaps the preferred alternative, is to run this loop only once but perform all reconstruction operations at each step. The reason this is currently impossible is the fact that scrambling does not (and cannot) use this mechanism but must be ordered in the exact way illustrated by the execute sequence function.³⁵

7.2.3 *Chain creation (all chains)*

Once the repair function detects that chain formation could be needed, a *create_chain* function in the phrase structure class is called. This function creates head-, A- and \bar{A} -chains, but does not (yet) apply to scrambling chains.

```
def create_chain(X):
    head, target = PhraseStructure.transfer_operation['prepare'](X)
    head.form_chain(target)
    if PhraseStructure.transfer_operation['test integrity'](target):
        target.create_chain() # Recursion, successive-cyclicity
```

If the reconstructed item needs further chain repair, the same function is called recursively. A complex head that hosts several affixes gets cleaned up completely, since only the last step creates a primitive head. In the case of phrasal targets, the result is successive-cyclic reconstruction. In other words, successive-cyclic reconstruction/movement is produced by the last four lines in this function.

The function is called by the probe head H (X).

³⁵ The most likely solution to this problem is to dissolve the scrambling operation into two separate operations, one taking care of the (i) right edge (unified with extraposition) and another which uses the (ii) bottom-up cycle. Then we can execute (i) as part of the extraposition step while (ii) gets unified with the rest of the mechanism.

```

def prepare_head_chain(X):
    return X, X.right().copy_for_chain()

def prepare_phrasal_chain(X):
    if X.A_bar_operator():
        X.prepare_A_bar_chain(X.container())
    else:
        PhraseStructure.transfer_operation = PhraseStructure.instructions['A-chain']      # Last Resort option
    return X.container(), X.copy_for_chain()

def prepare_A_bar_chain(X, probe):
    if not X == probe.next(probe.edge):
        probe = X.project_phonologically_null_head()
    probe.copy_criterial_features(X)
    probe.features = PhraseStructure.speaker_model.lexicon.apply_redundancy_rules(probe.features)

```

The chain is then prepared (see below) and created by *form_chain* function for the *target* element designated by the iteration.

```

def form_chain(X, target):
    for head in X.minimal_search_domain().minimal_search(PhraseStructure.transfer_operation['selection'],
                                                          PhraseStructure.transfer_operation['sustain']):
        if head != X and head.test_merge(target, PhraseStructure.transfer_operation['legible'], 'left'):
            break
        target.remove()
    else:
        if not X.top().bottom().test_merge(target, PhraseStructure.transfer_operation['legible'], 'right'):
            target.remove()
            if X.sister() and not target.scan_criterial_features('ΔOP'):
                X.sister().Merge_inside(target, 'left')
            else:
                X.top().bottom().Merge_inside(target, 'right')  # Last Resort option

```

Chain formation executes a minimal search (7.1.7) by using the selection and sustain parameters provided in the instructions and finds the closest position where the element is legible. If minimal search reaches the end but no suitable position came up, two further operations are attempted: merge to the right of the bottom node and then merge just below the probe head. The last option constitutes a last resort strategy that is always applied if nothing else works. Before the chain is formed, the operation is prepared (function *prepare_chain*).

```

def prepare_head_chain(X):
    return X, X.right().copy_for_chain()

def prepare_phrasal_chain(X):
    if X.A_bar_operator():
        X.prepare_A_bar_chain(X.container())
    else:
        PhraseStructure.transfer_operation = PhraseStructure.instructions['A-chain']    # Last Resort option
    return X.container(), X.copy_for_chain()

def prepare_A_bar_chain(X, probe):
    if not X == probe.next(probe.edge):
        probe = X.project_phonologically_null_head()
    probe.copy_criterial_features(X)
    probe.features = PhraseStructure.speaker_model.lexicon.apply_redundancy_rules(probe.features)

```

In the case of head chains, only the last line is executed which copies the target head before reconstruction (*copy_for_chain*; the *babtize* function provides chain indexes for the output). If the target element is a phrase, a decision is made between A/Ā-chain generation such that the latter is applied if and only if the reconstructed target phrase is an operator. A-chain is currently a last resort operation implemented by providing the default values for the instructions dictionary. See 7.2.5. If an Ā-chain is generated, then the feature content of the probe head is modified and a new probe head is generated in the case of double specifier structure. Once the chain has been prepared, it is formed.

7.2.4 Head chain

A head chain is triggered when the reconstruction cycle finds a complex head X(Y). Head Y is targeted inside X and moved downstream by minimal search (7.1.7) until a position is found where it can be selected. The chain creation algorithm is called recursively if Y itself is a complex head. Proper selection is defined as follows:

```

def properly_selected(X):
    return X.selector() and X.check_some(X.selector().licensed_complements())

```

Most likely this version does not yet handle long head movement that was part of previous models.

It has not been unified with the new architecture.

7.2.5 A-chains

A-chains constitute a last resort option when an operator \bar{A} -chain cannot be created and if the target satisfies conditions for the A-chain. The decision is currently based on whether the reconstructed specifier is an operator or not.

```
def prepare_phrasal_chain(X):
    if X.A_bar_operator():
        X.prepare_A_bar_chain(X.container())
    else:
        PhraseStructure.transfer_operation = PhraseStructure.instructions['A-chain']
    return X.container(), X.copy_for_chain()
```

The target phrase is merged to the first vacant position found by minimal search (7.1.7). A vacant position means either the complement of the bottom right node ($[XP_1\dots[Y __1]]$) or a position between two heads that does not contain other phrases ($[XP_1\dots[Y __1 [Z\dots]]]$).³⁶

```
def has_vacant_phrasal_position(X):
    return X.gapless_head() or X.is_right()
```

If the target element does not satisfy the conditions for A-chains, it is not reconstructed. It may be reconstructed by a later operation. Head reconstruction and scrambling reconstruction are applied before A-chains.

7.2.6 \bar{A} -chains

\bar{A} -chains are formed when the target element is an operator, i.e. contains an operator feature. Minimal search (7.1.7) finds the first specifier or complement position where the element can be selected.

³⁶ The operation is too simple to handle nontrivial A-chain data. Either an additional condition is added which requires that the operations targets only thematic position (licensing long-distance A-chains) or we add successive-cyclicity by relying on the recursion used in connection with head chains.

```

def Abar_legible(X, target):
    # Conditions for accepting [SpecXP] as landing site
    if target == X.next(X.edge) and \
        len(X.edge()) == 1 and \
        X.specifier_match(target) and \
        X.specifier_sister().tail_match(X.specifier_sister(), 'left'):
        return True
    # Effect of [CompXP] as landing site
    if target.is_right():
        target.adjunct = False
    # Conditions for accepting [CompXP] as landing site
    if X.sister() == target:
        if X.complement_match(target) and not (target.is_left() and X.specifier_mismatch(target)):
            return True

```

Self is the head H under consideration and *y* is the reconstructed phrase. The first clause handles reconstruction into a specifier position, the latter to the complement position. A specifier reconstruction is accepted if the H accepts *y* as a specifier, there is only one specifier position (tucking in operation is disabled) and the reconstructed object satisfies the tail test. Complement reconstruction is accepted if H accepts *y* as its complement.

7.2.7 Agree

Agree targets probe heads with an uninterpretable φ_- (which are by definition predicates) and locates a goal which values its φ -features. In Python:

```

def AgreeLF(X):
    X.value_from_goal(X.get_goal())

```

Self represents the probe head, which gets features from the goal selected by the *get_goal()* function. The goal is located by minimal search § 4.8.7, 7.1.7. In Python:

```

def get_goal(X):
    return next(X.minimal_search_domain().minimal_search(lambda x: x.goal_selection(),
                                                       lambda x: not x.phase_head()), None)

```

The function *minimal_search_domain()* determines the starting point of the search. For example, if the probe does not have a sister, then the search begins from the probe itself. Minimal search targets the selected starting point and looks for referential heads at their base positions and stops at the phase head. The user can define these properties by changing the lambda functions representing the function parameters. If nothing is found, the function results the probe itself. If a goal is found, its φ -features are *valued* to α .

```

def value_from_goal(X, goal):
    if goal:
        log(f'\n\t\Agree({X}^o, {goal.head()})')
        if feature_licensing(goal.head().interpretable_phi_features(), X.phi_bundles() and X.Condition_on_agreement_and_EPP(goal):
            X.value(goal)
        else:
            log(f'failed.')
            X.features.add('*')

```

The function gets the phi-features from the goal from a predetermined set defined by *target_phi_features()*. They are stored into (*phi*, *phi_*) tuples where *phi* is the incoming φ-feature and *phi_* is an unvalued counterpart if any, which results in valuation.

```

def value(X, goal):
    log(f'valued features ')
    for phi in (x for x in goal.head().features if x.startswith('iPHI') and f'PHI:{x.split(":")[1]}:_' in X.features):
        log(f'{[phi[5:]]} ')
        X.features.discard(f'PHI:{phi.split(":")[1]}:_')
        X.features.add(f'{phi[1:]}')
    X.features.update({'oLF'})
    X.features.add(f'PHI:IDX:{goal.head().get_id()}')

```

7.2.8 Scrambling (*scrambling_reconstruction.py*)

7.2.9 Adjunct promotion (*adjunct_constructor.py*)

Adjunct promotion is an operation that changes the status of a phrase structure from non-adjunct into an adjunct, thus it transfers the constituent from the “primary working space” into the “secondary working space.” Decisions concerning what will be an adjunct and non-adjunct cannot be made in tandem with consuming words from the input. The operation is part of transfer. If the phrase targeted by the operation is complex, the operation is trivial: the whole phrase structure is moved. If the targeted item is a primitive head, then there is an extra concern: how much of the surrounding structure should come with the head?

```

def externalize_structure(X):
    if X and X.head().is_adjoinable() and X.mother():
        if X.complex():
            X.externalize_and_transfer()
        else:
            X.externalize_head()

```

If the externalized element was a head, then we need to make a decision on whether its specifier should be taken as well. The decision is made as follows:

```

def externalize_head(X):
    if X.isolated_preposition():
        X.externalize_and_transfer()
    elif X.externalize_with_specifier():
        X.mother().mother().externalize_and_transfer()
    else:
        X.mother().externalize_and_transfer()

```

Specifier is carried if it is required by an EPP-feature or the special “capture specifier rule” applies, which is expressed by the latter function. The externalization function pulls the phrase into the secondary working space, adds tails features if needed, and transfers it:

```

def externalize_and_transfer(X):
    if X.mother():
        X.adjunct = True
        X.add_tail_features_if_missing()
        X.transfer_adjunct()

```

The transfer adjunct function detaches the adjunct temporarily from the structure before transfer is applied.

7.3 LF-legibility (LF.py)

7.3.1 Overall

The purpose of the LF-legibility test is to check that output of the syntactic pathway satisfies the LF-interface conditions and can therefore be interpreted semantically, at least in principle. Only primitive heads will be checked. The test consists of several independent tests, which are collected into a separate data structure as functions.

```

[('Edge feature test', PhraseStructure.edge_feature_tests),
 ('Selection test', self.selection_test),
 ('Projection Principle', PhraseStructure.projection_principle_failure),
 ('Head Integrity test', PhraseStructure.unrecognized_label),
 ('Feature Conflict test', PhraseStructure.feature_conflict),
 ('Probe-Goal test', PhraseStructure.probe_goal_test),
 ('Semantic Complement test', PhraseStructure.semantic_complement),
 ('Double Specifier Filter', PhraseStructure.double_spec_filter),
 ('Criterial Feature test', PhraseStructure.legitimate_criterial_feature),
 ('Adjunct Interpretation test', PhraseStructure.interpretable_adjunct),
 ('External head merge test', PhraseStructure.Complex_Head_Integrity)]

```

The LF-legibility test function serves as a “gateway” which regulates the tests that are selected for active use, and then calls the recursive test function. If no test battery is given as an argument, then all tests will be used; if a test battery is provided, it will be used. The recursive test function explores all primitive lexical items recursively in the output representation and applies all active tests to it.

```
def pass_LF_legibility(self, ps, logging=True):
    self.failed_feature = ''
    if not ps.copied:
        if ps.zero_level():
            for (test_name, test_failure) in self.active_test_battery:
                if test_failure(ps):
                    if logging:
                        log(f'\n\t{ps} failed {test_name}')
                    if self.failed_feature:
                        log(f'for [{self.failed_feature}]')
                    self.error_report_for_external_callers = f'{ps} failed {test_name}.'
                    return False
        else:
            if not self.pass_LF_legibility(ps.left(), logging):
                return False
            if not self.pass_LF_legibility(ps.right(), logging):
                return False
    return True
```

Of special interest is the selection test, which examines if the specifier and complement selection tests are valid. Selection test functions are provided in a separate data structure of this class.

```
self.selectionViolationTest = {'1EDGE': PhraseStructure.selection__negative_one_edge,
                               '-SPEC': PhraseStructure.selection__negative_specifier,
                               '-COMP': PhraseStructure.selection__negative_complement,
                               '!COMP': PhraseStructure.selection__positive_obligatory_complement,
                               '+COMP': PhraseStructure.selection__positive_disjunctive_complement,
                               '!!': PhraseStructure.selection__positive_self_selection,
                               '！': PhraseStructure.selection__negative_self_selection,
                               '+': PhraseStructure.selection__partial_self_selection}
```

They are applied to any lexical item with the gateway feature (left feature in the above dictionary).

```
def selectionTest(self, probe):
    for f in sorted(probe.features):
        key, feature_set = self.formatSelectionFeature(f)
        if key and not self.selectionViolationTest[key](probe, feature_set):
            self.failed_feature = str(feature_set)
    return True # test failed
```

The function activates a selection test if the lexical item has the gate feature, as listed in the selection test list above. If the test fails (*not test(probe, lexical_feature)*).

7.3.2 Selection tests

The selection rules, which are self-explanatory, are shown below.

```

def selection__negative_specifier(X, selected_feature_set):
    return not X.next(X.edge, lambda x: x.check(selected_feature_set) and not x.adjunct)

# Feature !1EDGE
def selection__negative_one_edge(X, selected_feature):
    return len(X.edge()) < 2

# Feature !COMP:L
def selection__positive_obligatory_complement(X, selected_feature_set):
    return X.selected_sister() and X.selected_sister().check(selected_feature_set)

# Feature +COMP:L,K
def selection__positive_disjunctive_complement(X, selected_feature_set):
    return X.selected_sister() and X.selected_sister().check_some(selected_feature_set)

# Feature -COMP:L
def selection__negative_complement(X, selected_feature_set):
    return not (X.proper_selected_complement() and X.proper_selected_complement().check_some(selected_feature_set))

# Feature []
def selection__positive_self_selection(X, selected_features):
    return X.check(set(selected_features.split(',')))

# Feature [-]
def selection__negative_self_selection(X, selected_features):
    return not X.check(set(selected_features.split(',')))

# Feature [+]
def selection__partial_self_selection(X, selected_features):
    return X.check_some(set(selected_features.split(',')))

```

7.3.3 Projection principle

Projection principle examines that all referential arguments receive a thematic role. The function first examines if the projection principle applies to the head α , and then verifies that the αP is contained inside a β which assigns it a thematic role.

```

def projection_principle_failure(X):
    return (X.max().projection_principle_applies() and
           not X.max().container_assigns_theta_role() or \
           X.projection_principleViolation())

```

The latter function is defined as follows:

```

def container_assigns_theta_role(X):
    Y = X.max().container()
    if Y:
        if Y.sister() == X or \
            (Y.geometrical_sister() == X and
             X.check_some(Y.licensed_complements())):
            return True
        return not Y.EF() and \
               Y.check_some({'SPEC: $\phi$ ', '!SPEC: $\phi$ '}) and \
               X.is_licensedSpecifier() and \
               not (X.selector() and X.selector().check({'-ARG'}))

```

The first condition (line 365) says that theta roles can be assigned via sisterhood $\{\alpha, \beta\}$. The second condition (line 366) examines the specifier-head configuration $[\beta_P \dots \alpha_P [\dots \beta \dots]]$.

7.4 Semantics (narrow_semantics.py)

7.4.1 Introduction

Semantic interpretation is implemented in the module *narrow_semantics.py* which bleeds the syntax-semantics interface (LF-interface). It begins by recursing through the whole structure and interpreting all lexical elements that have content understood by various semantic submodules.

```
def interpret_(self, ps):
    if not ps.copied:
        if ps.zero_level():
            if self.speaker_model.settings.retrieve('general_parameter_calculate_thematic_roles', True) and ps.theta_predicate():
                self.speaker_model.results.store_semantic_interpretation('Thematic roles', self.thematic_roles_module.reconstruct(ps))
            if self.speaker_model.settings.retrieve('general_parameter_calculate_predicates', True) and ps.check({'ARG'}): # and not ps.check({'\phi'}):
                self.speaker_model.results.store_semantic_interpretation('Predicates', self.predicates.reconstruct(ps))
            if self.speaker_model.settings.retrieve('UG_parameter_Agree', 'revised') == 'standard':
                self.predicates.operation_failed = False
        if ps.argument_by_agreement():
            self.speaker_model.results.store_semantic_interpretation('Indexing by Agree', self.predicates.reconstruct_agreement(ps))
        self.quantifiers_numerals_denotations_module.detect_phi_conflicts(ps)
        self.interpret_tail_features(ps)
        if self.speaker_model.settings.retrieve('general_parameter_project_objects', True):
            self.inventory_projection(ps)
        self.speaker_model.results.store_semantic_interpretation('Operator bindings', self.operator_variable_module.bind_operator(ps))
        self.speaker_model.results.store_semantic_interpretation('DIS-features', self.pragmatic_pathway.interpret_discourse_features(ps))
        if self.failure():
            return
    else:
        self.interpret_(ps.left())
        self.interpret_(ps.right())
```

All primitive lexical elements are targeted for interpretation; complex phrases are recursed. The general idea is that lexical features are diverged into different semantic subsystems for interpretation.

7.4.2 Projecting semantic inventories (semantic switchboard)

One function of narrow semantics is to project semantic objects, corresponding to the expressions in the input sentence, into the semantic inventories. This is done by function *inventory_projection()*. Suppose we are examining lexical item α . If α has lexical features that can be interpreted by one or several of the semantic subsystems, narrow semantics queries the corresponding system and, if that system can process that lexical feature, asks it to project the corresponding semantic object into existence and then links these objects to the original expression by using a lexical referential index feature. The referential index feature can be thought

of as a link between the expression and the corresponding semantic object. It has form [IDX:N,S] where N is a numerical identifier and S denotes the semantic space. Technically the query operation is implemented by using a router data structure *query* that channels lexical instructions to the subsystems that can process them. For example, command

```
if self.query[space] ['Accept'] (ps.head())
```

sends the instruction ‘Accept’ plus the head α to the semantic system given by the space parameter, which returns *True* if that system can accept and process α . The command

```
self.query[space] ['Project'] (ps, idx)
```

then asks the subsystem to project the corresponding element to the semantic inventory. Properties of the projected item are determined by the lexical features in α . Corresponding projection to the global inventory space will also take place.³⁷ The query data structure itself can be considered like a “switchboard” that is implemented as a dictionary of dictionaries, where the first level dictionary hosts the semantic space identifiers and the second the commands, and where the command is the key and the corresponding implementation function the value. The idea is that narrow semantics functions as a router that diverges the processing of lexical features to various cognitive subsystems.

7.4.3 *Quantifiers-numerals-denotations module*

The quantifiers-numerals-denotations (QND) module is specialized in processing referential expressions that involve “things” that can be quantified and counted. It projects semantic objects into its own semantic inventory that get linked with referential expressions occurring in phrase structure objects at the syntax-semantics interface. It can also understand and interpret referential lexical features such as φ -features and translate them into semantic features. The most important

³⁷ The assumption that every referential expression projects an entity into the discourse inventory might sound odd, since in many cases they should refer to an existing object instead. For example, pronoun *he* will typically denote an existing object. Indeed, it might. However, each time a referential expression is encountered in the input a new object is always projected, as shown by the code above, regardless of whether it will in the end be selected as a likely denotation

function of this module is the calculation of possible denotations and assignments. An assignment is intuitively a mapping between referential expressions in the sentence and denotations in the global discourse space.

7.4.4 Operator-variable interpretation (SEM_operators_variables.py)

The operator-variable module interprets operator-variable constructions. The kernel of the module is constituted by a function that binds operators [OP:F] with the finite propositional scope marker(s) {OP:F, FIN}. The follow function calculates operator bindings:

```
def bind_operator(self, head):
    self.bindings = []
    for operator_feature in (f for f in head.features if not self.scope_marker(head) and self.is_operator_feature(f)):
        binding = self.interpret_covert_scope(self.find_overt_scope(head, operator_feature))
        self.interpret_operator_variable_chain(binding, operator_feature)
    return self.bindings
```

Biding is determined by the result of overt scope computations and covert scope computations, which is then interpreted. The variable *binding* is a dictionary which contains information about the binding dependency. Overt scope computations are defined by

```
def find_overt_scope(head, operator_feature):
    return next(({'Head': head, 'Scope': scope, 'Overt': True} for scope in head.upward_path() if
               {operator_feature, 'Fin'}.issubset(scope.features)), {'Head': head, 'Scope': None, 'Overt': False})
```

which looks for a pair of operator and finiteness inside the working memory path. Covert scope is defined by

```
def interpret_covert_scope(binding):
    if not binding['Scope'] and '!SCOPE' not in binding['Head'].features:
        return next(({'Head': binding['Head'], 'Scope': scope, 'Overt': False}
                    for scope in binding['Head'].upward_path() if
                    scope.finite_left_periphery()),
                    {'Head': binding['Head'], 'Scope': None, 'Overt': False})
    return binding
```

which is activated only if overt scope computations have not produced results. It binds the operator to the local finite T or C.

7.4.5 Pragmatic pathway

The pragmatic pathway or module is involved in inferring and computing semantic information that we intuitively associate with (narrow) pragmatics. It has to do with propositional relations between thinkers (speaker, hearer) and propositions and their underlying communicative intentions. The system operates in two ways. On one hand it uses the incoming linguistic information and the context as a source material to infer pragmatic information, such as what is the topic and focus, and what type of communicative moves are involved. These inferential operations run silently in the background and do not change or alter the course of processing inside the syntactic pathway. The pragmatic pathway accesses the information through syntax-pragmatics interfaces. Secondly, the language system and the lexicon can grammaticalize features that activate operations inside the pragmatic pathway in a more direct way, which creates situations where grammatical devices (suffixes, words, prosody, word order, heads, lexical features) affect the pragmatic interpretation in a more direct way. This second mechanism operates through narrow syntax which routes discourse features to the pragmatic system for interpretation. Because these discourse features exist inside the syntactic pathway, they may affect syntactic processing as well.

Let us consider grammaticalized discourse features first. Narrow semantics has a function that directs discourse features to the pragmatic pathway for processing. This function will handle all discourse features. The operation sends each discourse feature for interpretation and then stores the results for later use. In the current version, discourse feature interpretation is merely registered in the output.

Calculations involved with the information structure are more fully developed. The operation is called during global semantic interpretation.

```
def calculate_information_structure(self, ps):
    log('\n\t\tCalculating information structure...')
    return self.create_topic_gradient(self.collect_arguments(ps))
```

First, the system determines what the root proposition is and what arguments we need to include into the information structural calculations. We are only interested in the thinker (speaker), proposition and the propositional attitude between the two. The system uses this information to calculate a *topic gradient*, which expresses what it thinks constitutes new and old information in the sentence being processed. The system works as follows. When new expressions are streamed into syntax, they are allocated attentional resources by the pragmatic system in the order they appear. This constitutes a syntax-pragmatics interface: it sends information from the syntactic pathway to the pragmatic system. It is registered and processed in the pragmatic system:

```
def allocate_attention(self, head):
    if head.referential() or head.preposition():
        idx = self.consume_index()
        head.features.add('*IDX: '+str(idx))
        self.records_of_attentional_processing[str(idx)] = {'Constituent': head.max(), 'Order': idx, 'Name': f'{head}'}
```

The first line determines what kind of elements are included into the attentional mechanism, and depends on the interests of the researcher. Then, some attentional resources are allocated to the processing, and order information is stored. Finally, when the sentence comes through the LF-interface, the pragmatic module attempts to construct the topic gradient on the basis of this and other sources of information. The presentation order at which linguistic information was originally presented and processed, as shown above, is now interpreted as representing relative topicality.³⁸

Topic gradient calculations are nontrivial for several reasons. First, they must ignore some noncanonical word order changes, such as those created by Ä dependencies. An interrogative direct object pronoun that occurs at the beginning of the sentence should not be interpreted as the topic. Second, the more noncanonical the position is, the more prominent the topic/focus interpretation tends to be. This is especially clear in Finnish. Third, this language allows one the

³⁸ It follows from this that noncanonical word orders can change the way expressions and the semantic objects are represented in the topic gradient. For example, in a language like Finnish with a relatively free word order, various word orders are correlated with distinct information structural interpretations, in fact to such an extent that some movement operations are called “topicalization” and “focussing.” This derives the discourse-configurationality profile.

topicalize/focus several constituents (multi-topic/focus constructions) and to perform sentence-internal topicalization/focusing.

Let us consider then the function that constructs the actual topic gradient when the whole sentence is interpreted. It takes the “constituents in information structure” as an argument, which lists the arguments that are part of the proposition the speaker has established a propositional attitude relation. The data structure `topic_gradient` is then built, which sorts the arguments/semantic objects under consideration and all information about them as recorded by the pragmatic module, and as ordered by their appearance in the comprehension process.

```
def create_topic_gradient(self, constituents_in_information_structure):
    marked_topic_lst = []
    topic_lst = []
    marked_focus_lst = []
    topic_gradient = {key: val for key, val in sorted(self.records_of_attentional_processing.items(), key=lambda ele: ele[0])}
    for key in topic_gradient:
        if topic_gradient[key]['Constituent'] in constituents_in_information_structure:
            if 'Marked gradient' in topic_gradient[key]:
                if topic_gradient[key]['Marked gradient'] == 'High':
                    marked_topic_lst.append(topic_gradient[key]['Name'])
                elif topic_gradient[key]['Marked gradient'] == 'Low':
                    marked_focus_lst.append(topic_gradient[key]['Name'])
            else:
                topic_lst.append(topic_gradient[key]['Name'])
    return {'Marked topics': marked_topic_lst, 'Neutral gradient': topic_lst, 'Marked focus': marked_focus_lst}
```

Next, the elements in the topic gradient are sorted into three lists “marked topics,” “neutral gradient” and “marked focus,” again in the order they appear in the `topic_gradient` data structure. The distribution is triggered by information that was stored in connection with the corresponding objects, here during transfer. Thus, in the function implementing adjunct reconstruction there is a function call which registers unexpected word orders used later in the creation of the three-tiered topic gradient. This creates another syntax-pragmatics interface mechanism. How this is done, and where this interface should be positioned in the general architecture, turned out to be extremely nontrivial problem and must be examined in the light of empirical data. The marked topic list, neutral gradient and marked focus lists then appear in the results of the simulation.

7.4.6 Argument-predicate pairs

Every predicate head α must be paired with an argument (or, in some cases, the argument will be assigned by default). The argument must, furthermore, be at the edge of the predicate α at the LF-

interface, where “edge” refers to referential heads that are merged “around” α : (i) features of α (pro), then (ii) the complement [α , XP], then (iii) the second-merge specifier [$_{\alpha P}$ XP [α YP]] and, if all these fail, (iv) an element from the extended edge which allows examination of objects merged outside of αP but part of α ’s upward path. The latter option results in what is usually referred to as finite or nonfinite control in the literature. The edge is searched for an argument in this specific order.

```
def identify_argument(X):
    available_arguments = [acquire(X) for acquire in [lambda x: x.pro_argument(),
                                                       lambda x: x.argument_complement(),
                                                       lambda x: x.local_edge(),
                                                       lambda x: x.control()]]
    return next((argument for argument in available_arguments if argument), None)
```

8 Formalization of the parser

8.1 Linear phase parser (`linear_phase_parser.py`)

8.1.1 The speaker model

The linear phase (LP) parser module `linear_phase_parser.py` defines the behavior of the parser. It defines an idealized speaker model for a speaker of some language. Languages and their speakers differ from each other. These differences are represented in the lexicon, most importantly in the functional lexicon. Each time the linear phase parser is instantiated, language is provided as a parameter which then applies the language-specific lexical redundancy rules to all lexical elements that make up the speaker model. The main script creates a separate speaker model for speakers of any language present in the lexicon. These speaker models differ only in terms of the composition of (some) lexical items; the computational core remains the same. We imagine the brain model as a container that hosts all modules and their connections, as shown in Figure 32.

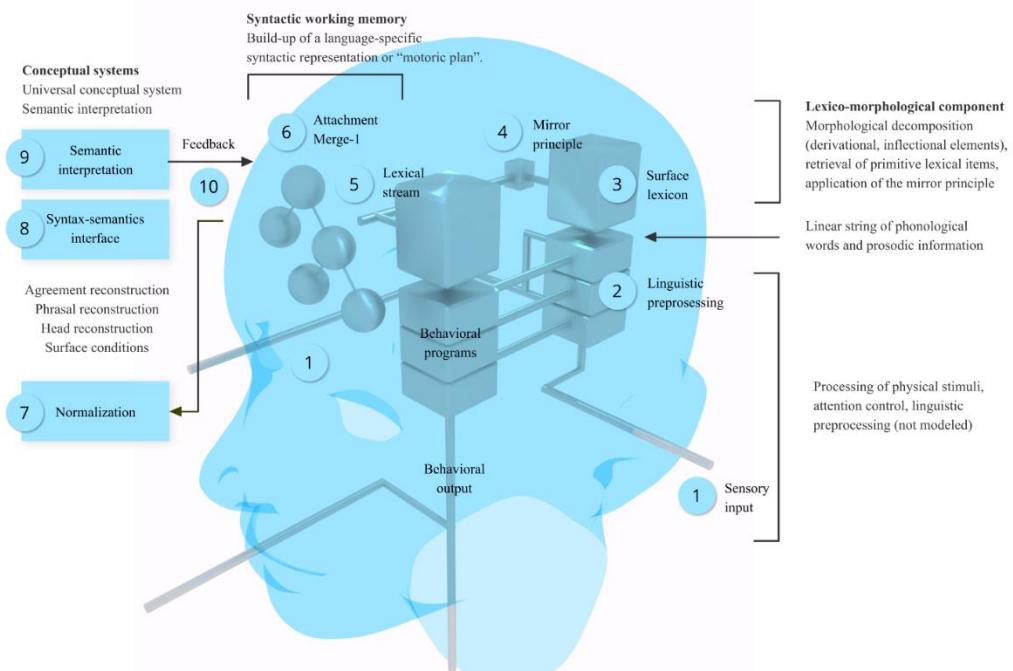


Figure 32. The speaker model contains all submodules defined by the theory and their connections.

8.1.2 Parse sentence (*parse_sentence*)

When the main script wants to parse a sentence, it sends the sentence to the speaker model (linear phase parser) as a list of words. The parser function prepares the parser by setting a host of input parameters (mostly having to do with logging and other support functions), and then calls for the recursive parser function *prase_new_item* with four arguments *current structure*, *list*, *index* and *inflection_buffer*, with *current structure* being empty, *index* = 0 and *inflection_buffer* empty.

```
def parse_sentence(self, index, lst):
    """Prepares the derivational search operation and
    calls the recursive derivational search function parse_new_item()"""

    # Bookkeeping and logging
    self.sentence = lst
    log_new_sentence(self, index, lst)
    PhraseStructure.chain_index = 0
    PhraseStructure.node_identity = 0

    # Initialize the components (mostly bookkeeping and logging)
    self.initialize()
    self.results.initialize(lst)
    self.plausibility_metrics.initialize()
    self.narrow_semantics.initialize()

    # Record lower level sensory processing (reading, hearing) for each word
    for word in self.sentence:
        self.results.consume_resources('Sensory Processing', word)

    # Call the derivational search function
    self.parse_new_item(None, lst, 0)
```

This function is self-explanatory. The last function (line 92) starts the recursive parsing operation.

8.1.3 Recursive parsing function (*parse_new_item*)

The recursive parsing function takes the currently constructed phrase structure α , a linearly ordered list of words, an index in the list of words and an inflection buffer as its arguments. These objects therefore represent the contents of the syntactic working memory at any processing phase.

It will first check if there is any reason to terminate processing. Processing is terminated if there are no more words, or if a self-termination flag `self.exit` has been raised somewhere during the execution.

```
# Stop processing if there are no more words to consume
if not self.circuit_breaker(ps, lst, index):

def circuit_breaker(self, ps, lst, index):
    set_logging(True)
    if self.exit:
        return True

    # If there are no more words, the solution is evaluated
    # at LF-interface and postsyntactically
    if index == len(lst):
        self.embedding = 0
        log_instance.indent_level = self.embedding
        self.evaluate_complete_solution(ps)
        return True
```

If there are no more words, contents of the syntactic working memory are send out for interpretation. This function is *complete_processing*. Suppose, however, that a new word w was consumed. At this point each word is phonological string. To retrieve properties of w, the lexicon will be accessed.

```
# Retrieve lexical items on the basis of phonological input
retrieved_lexical_items = self.lexicon.lexical_retrieval(lst[index])
```

This operation corresponds to a stage in language comprehension in which an incoming sensory-based item is matched with a lexical entry in the surface vocabulary. If w is ambiguous, all corresponding lexical items will be returned and will be explored in some order. The ordered list will be added to the recursive loop as an additional layer.

```
# Examine all retrieved lexical items (ambiguity resolution)
for lex in retrieved_lexical_items:
```

The lexicon is a mapping from phonological surface strings (keys) into constituents. If the lexical entry is mapped into a morphological decomposition, the resulting constituent will “contain” the

decomposition and no other properties. It constitutes a morphological chunk that will not and cannot enter syntax in this form; we can imagine these chunks are containing pointers to other elements in the lexicon. If the lexical entry maps into a primitive lexical item, then the constituent, primitive lexical item, will contain the set of features as specified in its lexical entry. If the lexical entry maps into an inflectional morpheme, then it will again consist of a set of (inflectional) features, but these will be processed differently. The following figure illustrates the idea.

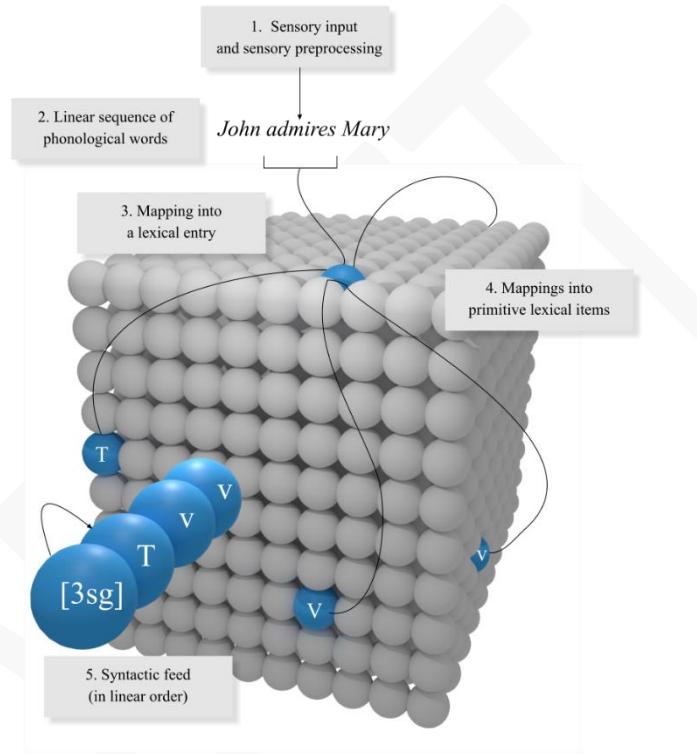


Figure 40. Organization of the lexicon. The lexicon is a mapping from phonological surface strings into constituents which make up the lexicon. Some surface strings map into primitive lexical items (blue items) which are sets of lexical features (e.g., /the/ ~ D). Other surface strings map into morphological chunks which contain pointers to further items (morphological decompositions)(e.g., /admires/ ~ V, v, T, 3sg).

We use the phrase structure class to generate lexical entries, whether they are morphologically complex or not. The intuitive motivation for this assumption is that at the bottom level the lexicon

has to map something into primitive lexical items; morphological chunks are just an additional layer build on the top of that foundation.

Next an item from this list of possible lexical entries will be subjected to *morphological parsing*.

Notice that a lexical item returned by the lexical retrieval may still consists of a morphological decomposition. The morphological parser will change the list of words that now contains the individual morphemes that were part of the original input word, in reverse order, together with the lexical item corresponding with the first item in the new list.

```
# Recursive branching: 1) morphological parsing, 2) inflection, 3) stream into syntax
# 1. Morphological parsing if applicable
if lex.morphological_chunk:
    self.embedding += 1
    self.parse_new_item(secure_copy(ps), lex.morphological_parse(ps, lst.copy(), index, inflection_buffer), index, inflection_buffer)

# 2. Process inflectional feature (withhold streaming to syntax)
if not lex.morphological_chunk and lex.inflectional:
    inflection_buffer = inflection_buffer | lex.features - {'inflectional'}
    log(f'= {fillu(lex.features)}')
    self.parse_new_item(secure_copy(ps), lst, index + 1, inflection_buffer)

# 3. Extract features from primitive lex and wrap them into primitive constituent and stream into syntax
if not lex.morphological_chunk and not lex.inflectional:
    self.explore_derivation_space(ps, self.lexical_stream.wrap(lex, inflection_buffer), lst, index)

inflection_buffer = set()
```

All word-internal morphemes are used to modify the original list of words, which now contains the morphological decomposition of the word in addition to the original phonological words. An input list *John + admires + Mary* will be transformed into *John + 3sg + T + v + admire + Mary* (ignoring the processing of the proper names). The *first item* in the new list will be sent to the syntactic component via lexical stream. The lexical stream pipeline handles several operations. Some of the morphemes could be inflectional, in which case they are stored as features into a separate inflectional memory buffer inside the lexical stream and then added to the next and hence also adjacent morpheme when it is being consumed in the reversed order. To allow backtracking from inflectional processing, the buffer is forwarded to the recursive parsing function as an input parameter. If inflectional features were encountered instead of a morpheme, the parsing function is called again immediately without any rank and merge operations. If there were several inflectional affixes, they would be all added to the next morpheme m consumed from the input. Other lexical items enter the syntactic module.

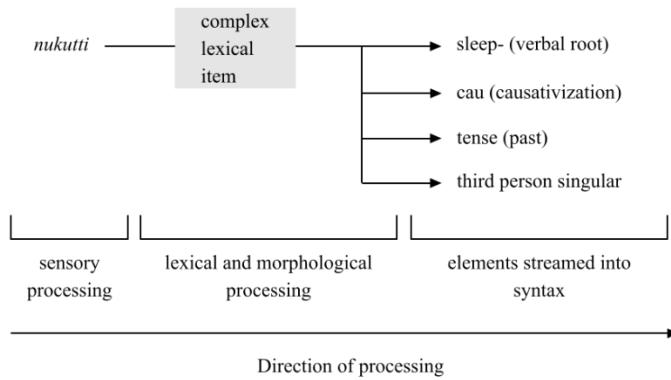


Figure 41. Lexical processing pipeline.

They will be merged to the existing phrase structure in the syntactic working memory, into some position. Merge sites that can be ruled out as impossible by using local information are filtered out. The remaining sites are ranked.

```
for N, transfer, address_label in self.plausibility_metrics.filter_and_rank(ps, X):
    new_constituent = ps.target_left_branch(N).attach(N, X, transfer, address_label)
    self.parse_new_item(new_constituent.top().copy(), lst, index + 1)
    if self.exit:
        break
```

Each site from the ranking is then explored. The loop examines each (site, transfer) pair provided by the plausibility function above. The candidate sites are right edge nodes in the partial phrase structure currently being developed. Transfer is a Boolean variable determining whether we want to transfer the left branch or not before Merge. The operation targets a node and attaches the incoming lexical item to it. The result will then be passed recursively to the parsing function.

The code above contains two essential functions: *target_left_branch* and *attach*. Recursive branching requires that we create a new structure when some solution is considered, which presupposes that we can identify equivalent nodes between these structures. The function *attach* performs Merge-1 and sinking depending on the situation and performs working memory operations.

```

def attach(X, site, terminal_lexical_item, transfer, address_label):
    if site.zero_level() and site.bottom_affix().word_internal():
        const = X.head_attachment(terminal_lexical_item)
    else:
        const = X.phrasal_attachment(terminal_lexical_item, transfer)
    X.speaker_model.results.consume_resources("Merge", const)
    log(f'{address_label}\n')
    return const

```

Once all words have been processed, the result will be submitted to a finalization stage implemented by function *evaluate_complete_solution*. This process will apply transfer, LF-legibility and semantic interpretation.

```

# Evaluates a complete solution at the LF-interface and semantic interpretation
def evaluate_complete_solution(self, ps):
    self.results.record_derivational_step(ps, 'PF-interface')
    log(f'\n\n\tPF/LF-interface mapping: -----')
    log(f'\n\t\t{ps}\n')
    ps.transfer_to_LF()
    ps = ps.top()
    log(f'\n\tLF-interface {ps}\n')
    self.results.record_derivational_step(ps, 'LF-interface')

    # Postsyntactic tests (LF-interface legibility and semantic interpretation)
    if self.postsyntactic_tests(ps):
        self.results.update_resources(PhraseStructure.resources, self.sentence)
        self.results.store_solution(ps)
        self.results.log_success(ps)
        self.results.record_derivational_step(ps, 'Accepted LF-interface')
        if self.settings.retrieve('general_parameter_only_first_solution', True):
            self.exit = True
    else:
        self.narrow_semantics.reset_for_new_interpretation()
        self.results.report_failure(ps)

    # Register one garden path if we evaluated complete solution but
    # there has not been any accepted solutions
    if not self.results.first_solution_found:
        self.results.consume_resources("Garden Paths", ps)

```

If these fail, the result is rejected; otherwise processing continues. Once a solution has been accepted and documented, control is returned to the parsing recursion.

8.2 Psycholinguistic plausibility

8.2.1 General architecture

The parser component obtains a list of possible attachment sites given some existing partial phrase structure α and an incoming lexical item β . This list is sent to the module `plausibility_metrics.py`

for processing. The parser will get a filtered and ranked list in return, which is used by the parser to explore solutions. The filter_and_rank function:

```
def filter_and_rank(self, ps, w):
    set_logging(False)
    if not w:
        return []
    if 'inflectional' in w.features:
        return []
    # Input integrity test
    if not ps or not w:
        return []
    elif ps.word_internal() and self.dispersion_filter_active():
        solutions = [(ps.bottom(), True, self.generate_address_label())]
    else:
        log(f'\n\tFiltering and ranking merge sites...')
        nodes_available = self.filter(ps.geometrical_minimal_search(), w)
        merge_sites = self.rank_merge_right_(nodes_available, w)
        solutions = self.evaluate_transfer(merge_sites)

    set_logging(True)
    log(f'\n\tRanking:')
    for i, (site, transfer, address_label) in enumerate(solutions, start=1):
        log('\n')
        if transfer:
            log("{:<80}{}".format(f'\t\t({{i}}) [{site}] + {{w.label()}}]', f'{address_label}'))
        else:
            log("{:<80}{}".format(f'\t\t({{i}}) [{site} + {{w.label()}}]', f'{address_label}'))
    return solutions
```

If the incoming word was part of the previous word, it will always be merged as its sister. We will therefore only return one solution. Otherwise, filter and ranking will be applied, in that order. Only nodes that passes the filter are ranked. Ranking uses cognitive parsing heuristics, as explained below. The user can activate and inactivate these heuristics in the configuration file.

8.2.2 Filtering

Filtering is implemented by going through all available nodes (i.e. those which are in the active working memory) and by rejecting them if a condition is satisfied. When a parsing branch is closed by filtering, it can never be explored by backtracking. The filtering conditions are the following: (1) The bottom node can be rejected if it has the property that it does not allow any complementizers. (2) Node α can be rejected if it constitutes a bad left branch (left branch filter). (3) $[\alpha \beta]$ can be rejected if it would break a configuration presupposed in word formation. (4) $[\alpha \beta]$ can be rejected if it constitutes an impossible sequence.

```

def filter(self, list_of_sites_in_active_working_memory, w):
    adjunction_sites = []
    for N in list_of_sites_in_active_working_memory:
        if N.does_not_accept_any_complements():
            log(f'Reject {N} + {w} because {N} does not accept complementizers. ')
            continue
        if N.complex() and self.left_branch_filter(N):
            log(f'Reject {N} + {w} due to bad left branch ({self.speaker_model.LF.error_report_for_external_callers})... ')
            continue
        if self.word_breaking_filter(N, w):
            log(f'Reject {N} + {w} because it breaks words. ')
            continue
        if N.impossible_sequence(w):
            log(f'Reject {N} + {w} because the sequence is impossible. ')
            continue
        adjunction_sites.append(N)
    return adjunction_sites

```

The left branch filter sends the phrase structure through the LF-interface and examines if transfer is successful. If it is not successful, the parsing path will be closed.

```

def left_branch_filter(self, N):
    dropped = N.copy().transfer_to_LF()
    self.speaker_model.LF.active_test_battery = self.left_branch_filter_test_battery
    left_branch_passes_LF = self.speaker_model.LF.pass_LF_legibility(dropped, False)
    if not left_branch_passes_LF:
        log(f'in {dropped}. ')
    return not left_branch_passes_LF

```

8.2.3 Ranking (*rank_merge_right_*)

Ranking forms a *baseline ranking* which it modifies by using various conditions. The baseline ranking is formed by `create_baseline_weighting()`.

```

def create_baseline_weighting(self, weighted_site_list, method=''):
    if method == 'Z':
        new_weighted_site_list = [(site, j) for j, (site, w) in enumerate(weighted_site_list, start=1)]
        new_weighted_site_list[0] = (new_weighted_site_list[0][0], len(weighted_site_list))
        new_weighted_site_list[-1] = (new_weighted_site_list[-1][0], len(weighted_site_list)+1)
        new_weighted_site_list = [(site, weight-1) for site, weight in new_weighted_site_list]
        return new_weighted_site_list
    if method == 'Random':
        site_w = list(range(len(weighted_site_list)))
        random.shuffle(site_w)
        return [(site, site_w[i]) for i, (site, w) in enumerate(weighted_site_list, start=0)]
    if method == 'Top-down':
        return [(site, -j) for j, (site, w) in enumerate(weighted_site_list, start=1)]
    if method == 'Bottom-up':
        return [(site, j) for j, (site, w) in enumerate(weighted_site_list, start=1)]
    if method == 'Sling':
        lst = [(site, -j) for j, (site, w) in enumerate(weighted_site_list, start=1)]      # Top-down
        lst[-1] = (lst[-1][0], 1)                # Promote the bottom node
        return lst
    else:
        return [(site, j) for j, (site, w) in enumerate(weighted_site_list, start=1)]

```

The weighted site list is a list of tuples (node, weight). Weights are initially formed from small numbers corresponding to the presupposed order, typically from 1 to number of nodes, but they could be anything. This order will be used if no further ranking is applied. Each (site, weight) pair

is examined and evaluated in the light of plausibility conditions which, when they apply, increase or decrease the weight provided for each site in the list. The function returns a list where the nodes have been ordered in decreasing order on the basis of its weights. Plausibility conditions are stored in a dictionary containing a pointer to the condition, weight, and logging information. Plausibility condition functions take α , β as input and evaluate whether they are true for this pair; if so, then the weight of α in the ranked list is modified according to the weight provided by the condition itself. In the current version the weight modifiers are ± 100 . These numbers outperform the small numbers assigned by the baseline weighting. They could compete on equal level if we assumed that the plausibility conditions provide smaller weight modifiers such as ± 1 . What the correct architecture is is an empirical matter that must be determined by psycholinguistic experimentation.

The following plausibility conditions are currently implemented. (1) Positive specifier selection: examines whether $[\alpha \beta]$ is supported by a position specifier selection feature for α at β . (2) Negative specifier selection: examines whether $[\alpha \beta]$ is rejected by a negative specifier selection feature for α at β . (3) Break head-complement relation: examines whether $[\alpha \beta]$ would break an existing head-complement selection. (4) Negative tail-test: examines whether $[\alpha \beta]$ would violate an internal tail-test at β . (5) Positive head-complement selection: examines if $[\alpha \beta]$ satisfies a complement selection feature of α for β . (6) Negative head-complement selection: examines if $[\alpha \beta]$ satisfies a negative complement selection feature of α for β . (7) Negative semantic match: examines $[\alpha \beta]$ violates a negative semantic feature requirement of α . (8) LF-legibility condition: examines if the left branch α in $[\alpha \beta]$ does not satisfy LF-legibility. (9) Negative adverbial test: examines if β has tail-features but does not satisfy the external tail-test. (10) Positive adverbial test: examines if β has tail-features and satisfies the external tail-test.

8.2.4 Knocking out heuristic principles

Heuristic principles can be knocked and/or controlled by *config_study.txt*. This is useful feature when we attempt to develop the principles, making it possible to observe their effects in isolation or in combination with other principles.

```
self.plausibility_conditions = \
    {'positive_spec_selection': \
        {'condition': self.positive_spec_selection,
         'weight': self.speaker_model.settings.retrieve('positive_spec_selection', 100),
         'log': '+Spec selection'},
     'negative_spec_selection': \
        {'condition': self.negative_spec_selection,
         'weight': self.speaker_model.settings.retrieve('negative_spec_selection', -100),
         'log': '-Spec selection'},
     'break_head_comp_relations': \
        {'condition': self.break_head_comp_relations,
         'weight': self.speaker_model.settings.retrieve('break_head_comp_relations', -100),
         'log': 'Head-complement word breaking condition'},
     'positive_head_comp_selection': \
        {'condition': self.positive_head_comp_selection,
         'weight': self.speaker_model.settings.retrieve('positive_head_comp_selection', 100),
         'log': '+Comp selection'},
     'negative_head_comp_selection': \
        {'condition': self.negative_head_comp_selection,
         'weight': self.speaker_model.settings.retrieve('negative_head_comp_selection', -100),
         'log': '-Comp selection'},
     'negative_semantics_match': \
        {'condition': self.negative_semantic_match,
         'weight': self.speaker_model.settings.retrieve('negative_semantics_match', -100),
         'log': 'Semantic mismatch'},
     'lf_legibility_condition': \
        {'condition': self.lf_legibility_condition,
         'weight': self.speaker_model.settings.retrieve('lf_legibility_condition', -100),
         'log': '-LF-legibility for left branch'},
     'negative_adverbial_test': \
        {'condition': self.negative_adverbial_test,
         'weight': self.speaker_model.settings.retrieve('negative_adverbial_test', -100),
         'log': '-Adverbial condition'},
     'positive_adverbial_test': \
        {'condition': self.positive_adverbial_test,
         'weight': self.speaker_model.settings.retrieve('positive_adverbial_test', 100),
         'log': '+Adverbial condition'}
    }
```

8.3 Resource consumption

The parser keeps a record of the computational resources consumed during the parsing of each input sentence. This allows the researcher to compare its operation to realistic online parsing processes acquired from experiments with native speakers.

The most important quantitative metric is the number of garden paths. It refers to the number of final but failed solutions evaluated at the LF-interface before an acceptable solution is found. If the number of 0, an acceptable solution was found during the first pass parse without backtracking. Number 1 means that the first pass parse failed, but the second solution was accepted, and so on. Notice that it only includes failed solutions after all words have been consumed. In a psycholinguistically plausible theory we should always get 0 expect in those cases in which native speakers too tend to arrive at failed solutions (as in *the horse raced past the barn fell*) at the end of consuming the input. The higher this number (>0) is, the longer it should take

native speakers to process the input sentence correctly (i.e. 1 = one failed solution, 2 = two failed solutions, and so on).

The number of various types of computational operations (e.g., Merge, Move, Agree) are also counted. Grammatical operations are counted as “black boxes” in the sense that we ignore all internal operations (e.g., minimal search, application of merge, generation of rejected solutions).

The number of head reconstructions, for example, is increased by one if and only if a head is moved from a starting position X into a final position Y; all intermediate positions and rejected solutions are ignored. This therefore quantifies the number of “standard” head reconstruction operations – how many times a head was reconstructed – that have been implemented during the processing of an input sentence. The number of all computational steps required to implement the said black box operation is always some linear function of that metric and is ignored. For example, countercyclic merge operations executed during head reconstruction will not show up in the number of merge operations; they are counted as being “inside” one successful head reconstruction operation. It is important to keep in mind, though, that each transfer operation will potentially increase the number independently of whether the solution was accepted or rejected.

For example, when the left branch α is evaluated during $[\alpha \beta]$, the operations are counted irrespective of whether α is rejected or accepted during the operation.

Counting is stopped after the first solution is found. This is because counting the number of operations consumed during an exhaustive search of solutions is psycholinguistically meaningless. It corresponds to an unnatural “off-line” search for alternative parses for a sentence that has been parsed successfully. This can be easily changed by the user, of course.

Resource counting is implemented by the parser and is recorded into a dictionary with keys referring to the type of operation (e.g., *Merge*, *Move Head*), value to the number of operations before the first solution was found.

```

self.resources = {"Total Time": {'ms': 0, 'n': 1},
                 "Garden Paths": {'ms': 1458, 'n': 0},
                 "Sensory Processing": {'ms': 75, 'n': 0},
                 "Lexical Retrieval": {'ms': 50, 'n': 0},
                 "Merge": {'ms': 7, 'n': 0},
                 "Head Chain": {'ms': 7, 'n': 0},
                 "Phrasal Chain": {'ms': 7, 'n': 0},
                 "Feature Inheritance": {'ms': 7, 'n': 0},
                 "Agree": {'ms': 7, 'n': 0},
                 "Feature": {'ms': 7, 'n': 0},
                 "Left Scrambling": {'ms': 7, 'n': 0},
                 "Right Scrambling": {'ms': 7, 'n': 0},
                 "Extraposition": {'ms': 7, 'n': 0},
                 "Last Resort Extraposition": {'ms': 7, 'n': 0},
                 "Mean time per word": {'ms': 0, 'n': 1}
}

```

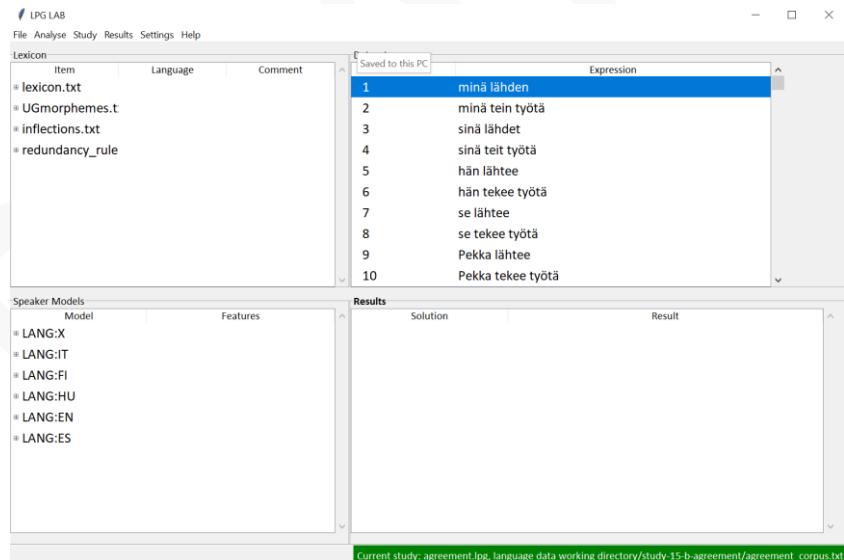
If the researcher adds more entries to this dictionary, they will show up in all resource reports.

The value is increased by function *consume_resources(key)* in the parser class. This function is called by procedures that successfully implement the computational operation (as determined by *key*), it increase the value by one unless the first solution has already been found. Thus, the user can add bookkeeping entries by adding the required key to the dictionary and then adding the line controlling_parsing_process.consume_resources("key") into the appropriate place in the code. For example, adding such entries to the phrase structure class would deliver resource consumption data from the lowest level (with a cost in processing speed). Resources are reported both in the results file and in a separate “_resources” file that is formatted so that it can be opened and analyzed easily with external programs, such as MS Excel. Execution time is reported in milliseconds. In Window the accuracy of this metric is ±15ms due to the way the operation system works. A simulation with 160 relatively basic grammatical sentences with the version of the program currently available resulted in 77ms mean processing time varying from <15ms to 265ms for sentences that exhibited no garden paths and 406ms for one sentence that involved 5 garden paths and hence severe difficulties in parsing.

9 Graphical user interface

9.1 Launching the interface and loading studies

The system can be used via a separate graphical user interface (GUI).³⁹ The interface can be launched by command `python lpparse -app` in the command prompt. Instead of running one study, the command opens up the following interface:



Note that as this component is being developed, the actual screen may differ from what is shown here. The program will first attempt to locate the application configuration file `$app_settings.txt`

³⁹ The interface was originally designed to replace the previous phrase structure image drawing algorithm that relied on the pyglet library that was not part of the standard Python installation. In addition, the current image generation component is based on *tkinter*, a library that can be used to design proper graphical user interfaces and therefore also more sophisticated drawing tools. The tree drawing algorithm in turn was originally designed to illustrate the output of the algorithm (it replicates what is in the output and does not allow the user to add or remove anything) and still serves this function, however as the graphical interface component expanded in scope it now offers several additional tools for the examination of the output of the model. This is what this interface is still primarily designed for, thus as it stands now it is not yet possible to replace all work with the command line tools with this interface although this may be an option in the future.

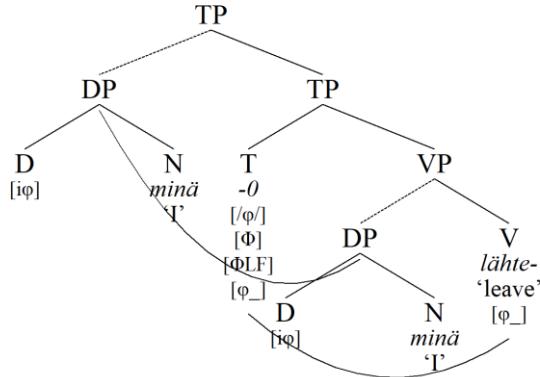
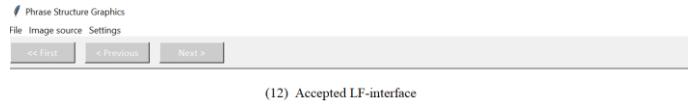
from the root directory and, if it is found, reads the name and folder of the study configuration file (*.lpg). The interface is then populated by the data and other properties as defined in the study configuration file. The current study configuration file and the corpus file as shown in the green bar at the bottom of the screen. We can load another study by selecting **File > Load Study...** from the main menu and then by selecting another study configuration file (file ending with lpg). Older study configuration files are not necessarily backwards compatible with the current version of the program. To replicate an older study, the preferred method is to download the frozen branch corresponding to that study from the source code repository. The user can save study configuration files by **File > Save Study...**

9.2 Analyzing individual sentences

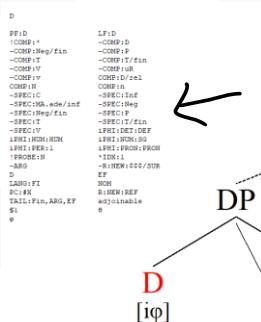
The main window opens with the current dataset as defined by the study configuration file.

Dataset	
Item	Expression
1	minä lähdēn
2	minä tein työtä
3	sinä lähdet
4	sinä teit työtä
5	hän lähtee
6	hän tekee työtä
7	se lähtee
8	se tekee työtä
9	Pekka lähtee
10	Pekka tekee työtä

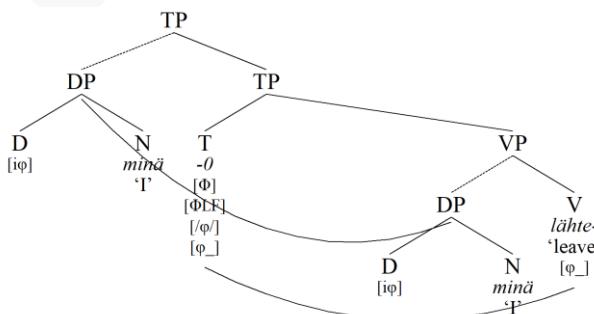
The user can then double click on any item in the list, after which the program will attempt to derive only that expression (also **Analyze > Analyze expression** in the main menu). If the derivation does not succeed and the sentence is judged ungrammatical, a window will pop up which shows the derivational log file (also **Results > examine log** from the main menu). Information concerning the reasons for failure are highlighted in red. If the derivation succeeds, a separate window opens which shows the first LF-interface structure derived for that expression:



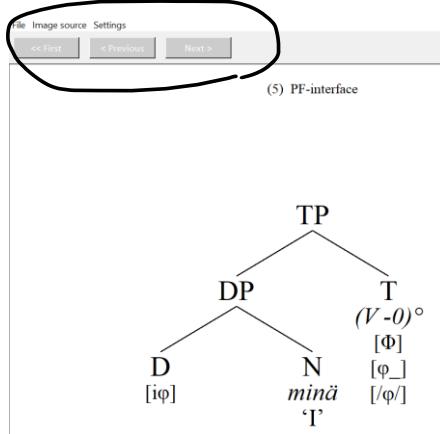
By moving the mouse cursor over any of the nodes in this phrase structure provides information about that node, currently its feature content:



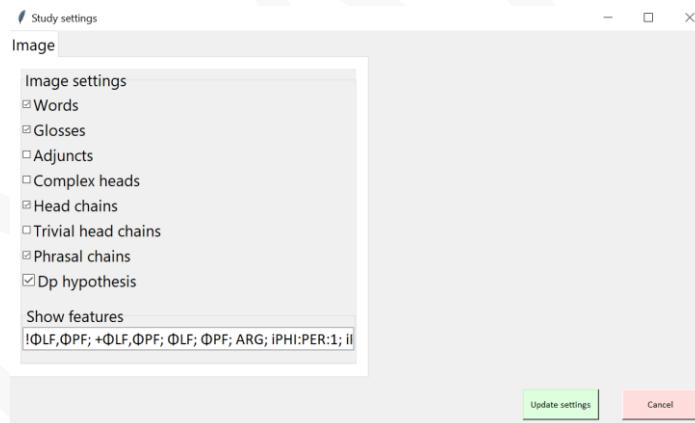
If the node is clicked once, it is possible to use the cursor buttons to move it in the phrase structure tree. For example:



This is handy if the image is cluttered. The window contains a few buttons that can be used to scroll the derivation both backwards and forward in time. For example, here we used the button **Previous** to find the PF-interface representation:



Finally, the menu item **Settings > Image settings...** allows the user to set various image parameters which determine what type of information is shown in the image.



The option **Dp hypothesis** replaces φPs with DPs for simplicity.

The results of the derivation are also visible in the bottom right block in the main window:



The user can then click on any of the elements and get further information. For example, clicking on **Resources** provides a list of resources consumed during the derivation:

Results	
Solution	Result
# 1	[<D minä>:3 [-0 [<__>:3 lähte-]]]
Ontology	
Resources	
Total Time	1 x 913 = 913 ms
Garden Paths	0 x 1458 = 0 ms
Sensory Processing	2 x 75 = 150 ms
Lexical Retrieval	14 x 50 = 700 ms
Merge	3 x 7 = 21 ms
Head Chain	2 x 7 = 14 ms
Phrasal Chain	0 x 7 = 0 ms

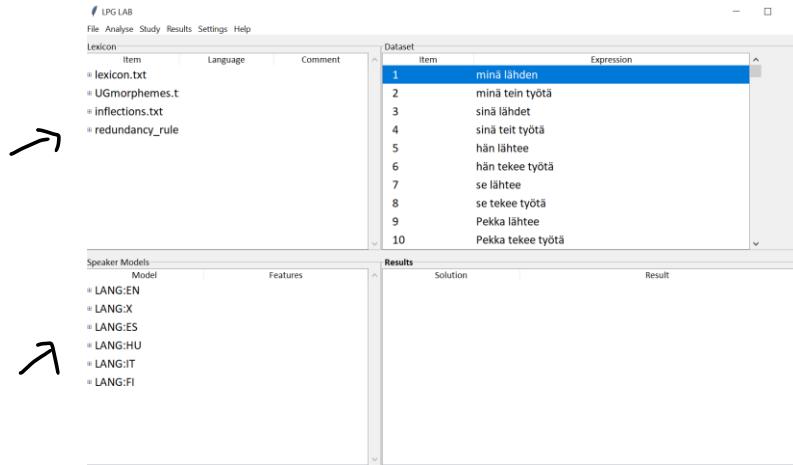
The notation X x Y means that the algorithm performed the operation X times, each which was set to consume Y milliseconds.

9.3 Running one complete study

The user can run one complete study by selecting **Study > Run complete study** from the main menu. This is equivalent to writing *python lpparse* into the command prompt.

9.4 Examining the root lexicons and speaker models

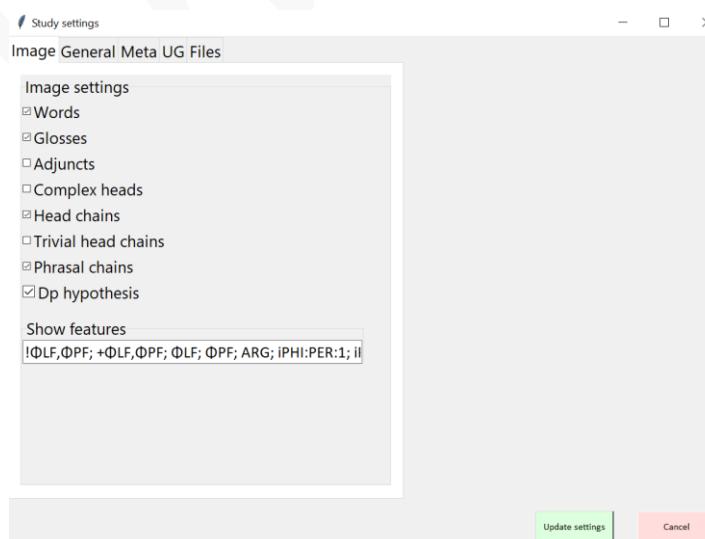
Information about the root lexicons and speaker models is provided in the two panels at the left in the main window:



By clicking on the small + we get more information about what each of these objects contain. The upper left panel contains the lexical entries as they were extracted from the files named here. The lower left panel contains information about the speaker models, in particular, about their lexicons. This information is based on the lexical and lexical redundancy files as specified in the study configuration file.

9.5 Simulation parameters and settings

Most simulation parameters defined in the study configuration file can be configured also from the main menu **Settings > Settings...** which then opens up a dialog window where the various settings can be defined:



The various settings are available in the separate tabs. They correspond to the settings available in the study configuration file, but currently not all parameters available in the latter are not present in the former. Once the settings have been changed, they can be saved by clicking **Update settings** and the **File > Save Study As....** The result is the same as if the changes were made directly into the study configuration file.

DRAFT

References

DRAFT