

Chapter 4: Features and Constraints

At the end of the class you should be able to:

- recognize and represent constraint satisfaction problems
- show how constraint satisfaction problems can be solved with search
- implement and trace arc-consistency of a constraint graph
- show how domain splitting can solve constraint problems

Posing a Constraint Satisfaction Problem

A CSP is characterized by

- A set of variables V_1, V_2, \dots, V_n .
- Each variable V_i has an associated domain \mathbf{D}_{V_i} of possible values.
- There are hard constraints on various subsets of the variables which specify legal combinations of values for these variables.
- A solution to the CSP is an assignment of a value to each variable that satisfies all the constraints.

Example: scheduling activities

- **Variables:** A, B, C, D, E that represent the starting times of various activities.
- **Domains:** $\mathbf{D}_A = \{1, 2, 3, 4\}$, $\mathbf{D}_B = \{1, 2, 3, 4\}$,
 $\mathbf{D}_C = \{1, 2, 3, 4\}$, $\mathbf{D}_D = \{1, 2, 3, 4\}$, $\mathbf{D}_E = \{1, 2, 3, 4\}$
- **Constraints:**

$$(B \neq 3) \wedge (C \neq 2) \wedge (A \neq B) \wedge (B \neq C) \wedge \\ (C < D) \wedge (A = D) \wedge (E < A) \wedge (E < B) \wedge \\ (E < C) \wedge (E < D) \wedge (B \neq D).$$

Generate-and-Test Algorithm

- Generate the assignment space $\mathbf{D} = \mathbf{D}_{V_1} \times \mathbf{D}_{V_2} \times \dots \times \mathbf{D}_{V_n}$.
Test each assignment with the constraints.

- Example:

$$\begin{aligned}\mathbf{D} &= \mathbf{D}_A \times \mathbf{D}_B \times \mathbf{D}_C \times \mathbf{D}_D \times \mathbf{D}_E \\ &= \{1, 2, 3, 4\} \times \{1, 2, 3, 4\} \times \{1, 2, 3, 4\} \\ &\quad \times \{1, 2, 3, 4\} \times \{1, 2, 3, 4\} \\ &= \{\langle 1, 1, 1, 1, 1 \rangle, \langle 1, 1, 1, 1, 2 \rangle, \dots, \langle 4, 4, 4, 4, 4 \rangle\}.\end{aligned}$$

- How many assignments need to be tested for n variables each with domain size d ?

Backtracking Algorithms

- Systematically explore \mathbf{D} by instantiating the variables one at a time
- evaluate each constraint predicate as soon as all its variables are bound
- any partial assignment that doesn't satisfy the constraint can be pruned.

Example Assignment $A = 1 \wedge B = 1$ is inconsistent with constraint $A \neq B$ regardless of the value of the other variables.

A CSP can be solved by graph-searching:

- A node is an assignment values to some of the variables.
- Suppose node N is the assignment $X_1 = v_1, \dots, X_k = v_k$.
Select a variable Y that isn't assigned in N .
For each value $y_i \in \text{dom}(Y)$
 $X_1 = v_1, \dots, X_k = v_k, Y = y_i$ is a neighbour if it is consistent with the constraints.
- The start node is the empty assignment.
- A goal node is a total assignment that satisfies the constraints.

Consistency Algorithms

- Idea: prune the domains as much as possible before selecting values from them.
- A variable is **domain consistent** if no value of the domain of the node is ruled impossible by any of the constraints.
- **Example:** Is the scheduling example domain consistent?

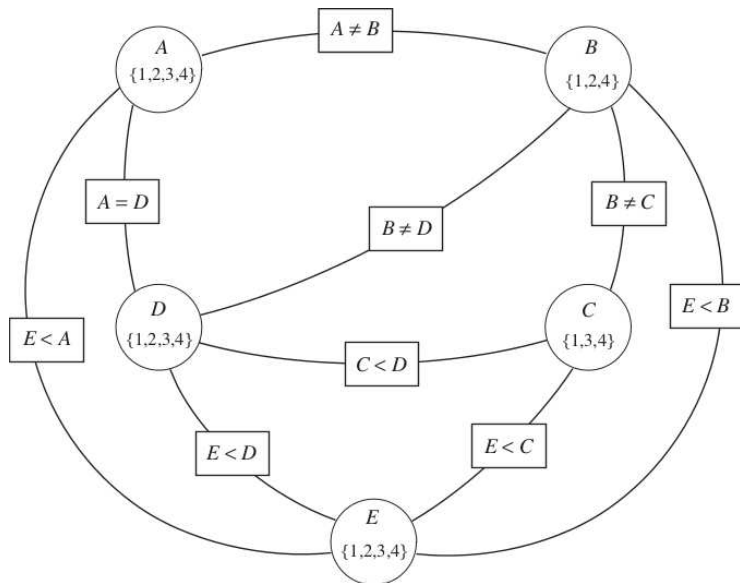
Consistency Algorithms

- Idea: prune the domains as much as possible before selecting values from them.
- A variable is **domain consistent** if no value of the domain of the node is ruled impossible by any of the constraints.
- **Example:** Is the scheduling example domain consistent?
 $D_B = \{1, 2, 3, 4\}$ isn't domain consistent as $B = 3$ violates the constraint $B \neq 3$.

Constraint Network

- There is a oval-shaped node for each variable.
- There is a rectangular node for each constraint.
- There is a domain of values associated with each variable node.
- There is an arc from variable X to each constraint that involves X .

Example Constraint Network



- An arc $\langle X, r(X, \bar{Y}) \rangle$ is **arc consistent** if, for each value $x \in \text{dom}(X)$, there is some value $\bar{y} \in \text{dom}(\bar{Y})$ such that $r(x, \bar{y})$ is satisfied.
- A network is arc consistent if all its arcs are arc consistent.
- What if arc $\langle X, r(X, \bar{Y}) \rangle$ is *not* arc consistent?

- An arc $\langle X, r(X, \bar{Y}) \rangle$ is **arc consistent** if, for each value $x \in \text{dom}(X)$, there is some value $\bar{y} \in \text{dom}(\bar{Y})$ such that $r(x, \bar{y})$ is satisfied.
- A network is arc consistent if all its arcs are arc consistent.
- What if arc $\langle X, r(X, \bar{Y}) \rangle$ is *not* arc consistent?
All values of X in $\text{dom}(X)$ for which there is no corresponding value in $\text{dom}(\bar{Y})$ can be deleted from $\text{dom}(X)$ to make the arc $\langle X, r(X, \bar{Y}) \rangle$ consistent.

Arc Consistency Algorithm

- The arcs can be considered in turn making each arc consistent.
- When an arc has been made arc consistent, does it ever need to be checked again?

Arc Consistency Algorithm

- The arcs can be considered in turn making each arc consistent.
- When an arc has been made arc consistent, does it ever need to be checked again?

An arc $\langle X, r(X, \overline{Y}) \rangle$ needs to be revisited if the domain of one of the Y 's is reduced.

- Three possible outcomes when all arcs are made arc consistent: (Is there a solution?)
 - ▶ One domain is empty \implies
 - ▶ Each domain has a single value \implies
 - ▶ Some domains have more than one value \implies

Arc Consistency Algorithm

- The arcs can be considered in turn making each arc consistent.
- When an arc has been made arc consistent, does it ever need to be checked again?

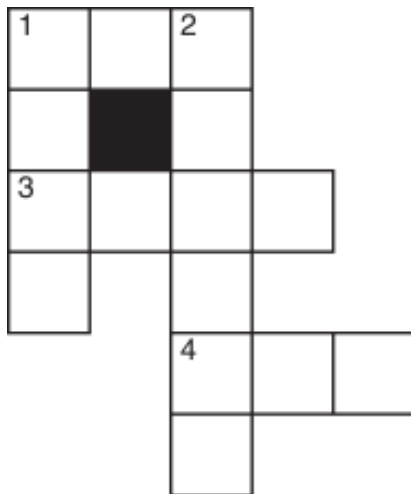
An arc $\langle X, r(X, \overline{Y}) \rangle$ needs to be revisited if the domain of one of the Y 's is reduced.

- Three possible outcomes when all arcs are made arc consistent: (Is there a solution?)
 - ▶ One domain is empty \implies no solution
 - ▶ Each domain has a single value \implies unique solution
 - ▶ Some domains have more than one value \implies there may or may not be a solution

Finding solutions when AC finishes

- If some domains have more than one element \implies search
- Split a domain, then recursively solve each half.
- It is often best to split a domain in half.
- Do we need to restart from scratch?

Example: Crossword Puzzle



Words:

ant, big, bus, car, has
book, buys, hold,
lane, year
beast, ginger, search,
symbol, syntax

Hard and Soft Constraints

- Given a set of variables, assign a value to each variable that either
 - ▶ satisfies some set of constraints: **satisfiability problems** — “hard constraints”
 - ▶ minimizes some cost function, where each assignment of values to variables has some cost: **optimization problems** — “soft constraints”
- Many problems are a mix of hard and soft constraints (called constrained optimization problems).

Local Search (Greedy Descent):

- Maintain an assignment of a value to each variable.
- Repeat:
 - ▶ Select a variable to change
 - ▶ Select a new value for that variable
- Until a satisfying assignment is found

- Aim: find an assignment with zero unsatisfied constraints.
- Given an assignment of a value to each variable, a **conflict** is an unsatisfied constraint.
- The goal is an assignment with zero conflicts.
- Heuristic function to be minimized: the number of conflicts.

To choose a variable to change and a new value for it:

- Find a variable-value pair that minimizes the number of conflicts
- Select a variable that participates in the most conflicts.
Select a value that minimizes the number of conflicts.
- Select a variable that appears in any conflict.
Select a value that minimizes the number of conflicts.
- Select a variable at random.
Select a value that minimizes the number of conflicts.
- Select a variable and value at random; accept this change if it doesn't increase the number of conflicts.

- When the domains are small or unordered, the neighbors of an assignment can correspond to choosing another value for one of the variables.
- When the domains are large and ordered, the neighbors of an assignment are the adjacent values for one of the variables.
- If the domains are continuous, **Gradient descent** changes each variable proportionally to the gradient of the heuristic function in that direction.

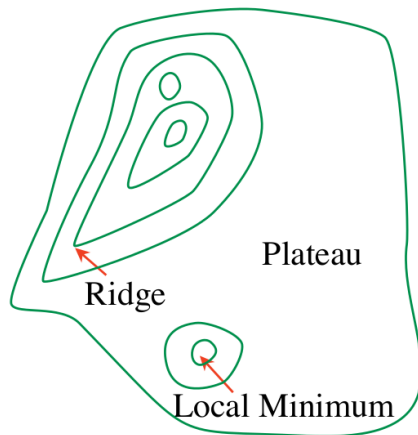
The value of variable X_i goes from v_i to

- When the domains are small or unordered, the neighbors of an assignment can correspond to choosing another value for one of the variables.
- When the domains are large and ordered, the neighbors of an assignment are the adjacent values for one of the variables.
- If the domains are continuous, **Gradient descent** changes each variable proportionally to the gradient of the heuristic function in that direction.

The value of variable X_i goes from v_i to $v_i - \eta \frac{\partial h}{\partial X_i}$.
 η is the step size.

Problems with Greedy Descent

- a local minimum that is not a global minimum
- a plateau where the heuristic values are uninformative
- a ridge is a local minimum where n -step look-ahead might help



- Consider two methods to find a minimum value:
 - ▶ Greedy descent, starting from some position, keep moving down & report minimum value found
 - ▶ Pick values at random & report minimum value found
- Which do you expect to work better to find a global minimum?
- Can a mix work better?

Randomized Greedy Descent

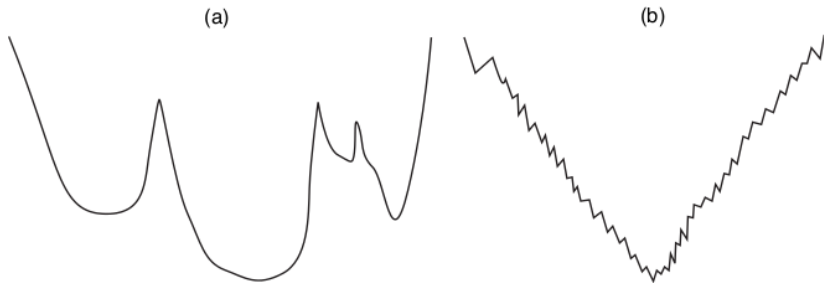
As well as downward steps we can allow for:

- **Random steps:** move to a random neighbor.
- **Random restart:** reassign random values to all variables.

Which is more expensive computationally?

1-Dimensional Ordered Examples

Two 1-dimensional search spaces; step right or left:



- Which method would most easily find the global minimum?
- What happens in hundreds or thousands of dimensions?
- What if different parts of the search space have different structure?

Stochastic local search is a mix of:

- Greedy descent: move to a lowest neighbor
- Random walk: taking some random steps
- Random restart: reassigning values to all variables

Variants of random walk:

- When choosing the best variable-value pair, randomly sometimes choose a random variable-value pair.
- When selecting a variable then a value:
 - ▶ Sometimes choose any variable that participates in the most conflicts.
 - ▶ Sometimes choose any variable that participates in any conflict (a red node).
 - ▶ Sometimes choose any variable.
- Sometimes choose the best value and sometimes choose a random value.

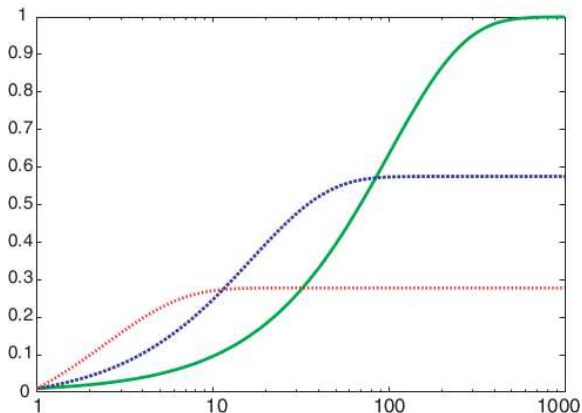
- How can you compare three algorithms when
 - ▶ one solves the problem 30% of the time very quickly but doesn't halt for the other 70% of the cases
 - ▶ one solves 60% of the cases reasonably quickly but doesn't solve the rest
 - ▶ one solves the problem in 100% of the cases, but slowly?

Comparing Stochastic Algorithms

- How can you compare three algorithms when
 - ▶ one solves the problem 30% of the time very quickly but doesn't halt for the other 70% of the cases
 - ▶ one solves 60% of the cases reasonably quickly but doesn't solve the rest
 - ▶ one solves the problem in 100% of the cases, but slowly?
- Summary statistics, such as mean run time, median run time, and mode run time don't make much sense.

Runtime Distribution

- Plots runtime (or number of steps) and the proportion (or number) of the runs that are solved within that runtime.



Variant: Simulated Annealing

- Pick a variable at random and a new value at random.
- If it is an improvement, adopt it.
- If it isn't an improvement, adopt it probabilistically depending on a temperature parameter, T .
 - ▶ With current assignment n and proposed assignment n' we move to n' with probability $e^{(h(n')-h(n))/T}$
- Temperature can be reduced.

Variant: Simulated Annealing

- Pick a variable at random and a new value at random.
- If it is an improvement, adopt it.
- If it isn't an improvement, adopt it probabilistically depending on a temperature parameter, T .
 - ▶ With current assignment n and proposed assignment n' we move to n' with probability $e^{(h(n')-h(n))/T}$
- Temperature can be reduced.

Probability of accepting a change:

Temperature	1-worse	2-worse	3-worse
10	0.91	0.81	0.74
1	0.37	0.14	0.05
0.25	0.02	0.0003	0.000006
0.1	0.00005	2×10^{-9}	9×10^{-14}

- To prevent cycling we can maintain a **tabu list** of the k last assignments.
- Don't allow an assignment that is already on the tabu list.
- If $k = 1$, we don't allow an assignment of to the same value to the variable chosen.
- We can implement it more efficiently than as a list of complete assignments.
- It can be expensive if k is large.

A total assignment is called an **individual**.

- **Idea:** maintain a population of k individuals instead of one.
- At every stage, update each individual in the population.
- Whenever an individual is a solution, it can be reported.
- Like k restarts, but uses k times the minimum number of steps.

- Like parallel search, with k individuals, but choose the k best out of all of the neighbors.
- When $k = 1$, it is greedy descent.
- When $k = \infty$, it is breadth-first search.
- The value of k lets us limit space and parallelism.

Stochastic Beam Search

- Like beam search, but it probabilistically chooses the k individuals at the next generation.
- The probability that a neighbor is chosen is proportional to its heuristic value.
- This maintains diversity amongst the individuals.
- The heuristic value reflects the fitness of the individual.
- Like asexual reproduction: each individual mutates and the fittest ones survive.

- Like stochastic beam search, but pairs of individuals are combined to create the offspring:
- For each generation:
 - ▶ Randomly choose pairs of individuals where the fittest individuals are more likely to be chosen.
 - ▶ For each pair, perform a cross-over: form two offspring each taking different parts of their parents:
 - ▶ Mutate some values.
- Stop when a solution is found.

- Given two individuals:

$$X_1 = a_1, X_2 = a_2, \dots, X_m = a_m$$

$$X_1 = b_1, X_2 = b_2, \dots, X_m = b_m$$

- Select i at random.
- Form two offspring:

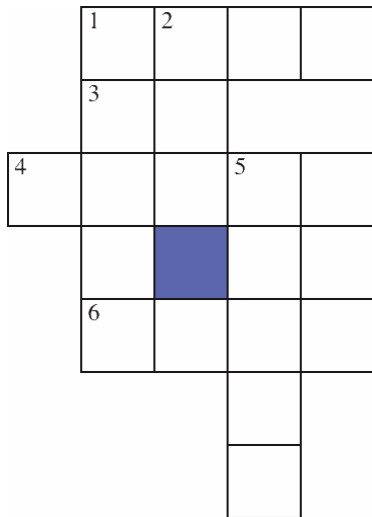
$$X_1 = a_1, \dots, X_i = a_i, X_{i+1} = b_{i+1}, \dots, X_m = b_m$$

$$X_1 = b_1, \dots, X_i = b_i, X_{i+1} = a_{i+1}, \dots, X_m = a_m$$

- The effectiveness depends on the ordering of the variables.
- Many variations are possible.

- A **Constraint Satisfaction problem** consists of:
 - ▶ a set of variables
 - ▶ a set of possible values, a **domain** for each variable
 - ▶ a set of constraints amongst subsets of the variables
- The aim is to find a set of assignments that satisfies all constraints, or to find all such assignments.

Example: crossword puzzle



at, be, he, it, on,
eta, hat, her, him,
one,
desk, dove, easy,
else, help, kind,
soon, this,
dance, first, fuels,
given, haste, loses,
sense, sound, think,
usage

Two ways to represent the crossword as a CSP

- First representation:
 - ▶ nodes represent word positions: 1-down...6-across
 - ▶ domains are the words
 - ▶ constraints specify that the letters on the intersections must be the same.
- Dual representation:
 - ▶ nodes represent the individual squares
 - ▶ domains are the letters
 - ▶ constraints specify that the words must fit

- First representation:
 - ▶ nodes represent the chains and regions
 - ▶ domains are the scene objects
 - ▶ constraints correspond to the intersections and adjacency
- Dual representation:
 - ▶ nodes represent the intersections
 - ▶ domains are the intersection labels
 - ▶ constraints specify that the chains must have same marking

- Agreement
- Linear Order and Optionality
- Structural Interpretation

- In many languages word forms carry morpho-syntactic features
- Usually morpho-syntactic features are highly ambiguous
- Determiner, adjectives and nouns in German noun phrases have to agree in case, number, and gender

	number	gender	case
die	sing, plur	masc, fem, neutr	nom, acc
großen	sing, plur	masc, fem, neutr	nom, gen, dat, acc
Teller	sing, plur	masc	nom, gen, dat, acc

- The possible value combinations are heavily constrained
- The constraints can be extensionally specified

<i>die</i>	⟨sing, fem, nom⟩	⟨sing, fem, acc⟩
	⟨plur, masc, nom⟩	⟨plur, masc, acc⟩
	⟨plur, fem, nom⟩	⟨plur, fem, acc⟩
	⟨plur, neutr, nom⟩	⟨plur, neutr, acc⟩
<i>großen</i>	⟨sing, masc, gen⟩	⟨sing, masc, dat⟩
	⟨sing, fem, gen⟩	⟨sing, fem, dat⟩
	⟨sing, neutr, gen⟩	⟨sing, neutr, dat⟩
	⟨plur, masc, nom⟩	⟨plur, masc, gen⟩
	⟨plur, masc, dat⟩	...
<i>Teller</i>	⟨sing, masc, nom⟩	⟨sing, masc, dat⟩
	⟨sing, masc, acc⟩	⟨plur, masc, nom⟩
	⟨plur, masc, gen⟩	⟨plur, masc, acc⟩

- ... or as a single logical expression

die $fem \wedge sing \wedge (nom \vee acc)$
 $\vee plur \wedge (masc \vee fem \vee neutr) \wedge (nom \vee acc)$

großen $sing \wedge (gen \vee dat) \wedge (masc \vee fem \vee neutr)$
 $\vee plur \wedge (masc \vee fem \vee neutr)$
 $\wedge (nom \vee gen \vee dat \vee acc)$

Teller $masc \wedge (sing \wedge (nom \vee dat \vee acc)$
 $\vee plur \wedge (nom \vee gen \vee acc)))$

- ... or as separate constraints

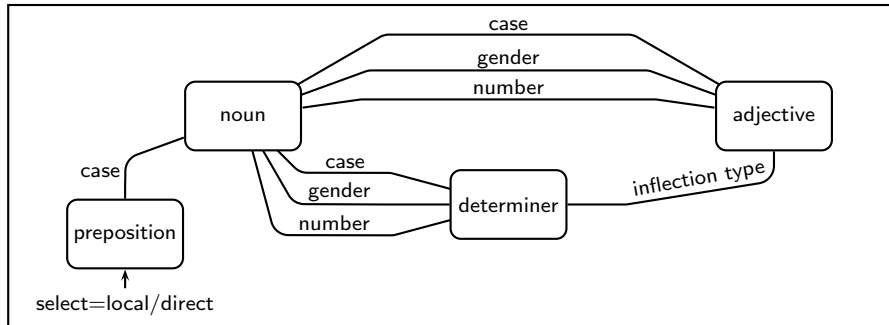
die masc \vee fem \vee neutr
 sing \vee plur
 nom \vee acc
 sing \rightarrow fem

großen nom \vee gen \vee dat \vee acc
 masc \vee fem \vee neutr
 sing \vee plur
 sing \wedge masc \rightarrow gen \vee dat \vee acc
 sing \wedge (fem \vee neutr) \rightarrow gen \vee dat

Teller masc
 sing \vee plur
 sing \rightarrow nom \vee dat \vee acc
 plur \rightarrow nom \vee gen \vee acc

- Agreement requires two or more word forms to share the same feature value
- Agreement is imposed in certain structural contexts, e.g. in German
 - ▶ noun phrases: determiner, adjective, noun
features: number, gender, case
der kluge Hund, des klugen Hundes, dem klugen Hund, den klugen Hund, die klugen Hunde, ...
 - ▶ clause-level: subject-verb(-reflexive pronoun):
features: person, number
Ich freue mich. Du freust dich. Er freut sich. ...
- Checking for agreement is a (simple) constraint satisfaction problem

Agreement



- Partial order, e.g. German prepositional phrase

- ▶ Examples

auf das Haus

auf das kleine Haus

auf das ziemlich kleine Haus

aufs Haus

- ▶ Constraints

Preposition < Determiner

Contracted Preposition < Graduating Particle

Determiner < Graduating Particle

Graduating Particle < Adjective

Adjective < Noun

- Co-occurrence constraints, e.g. German prepositional phrase

- ▶ Examples

auf das Haus

auf das kleine Haus

auf das ziemlich kleine Haus

aufs Haus

- ▶ Constraints

preposition \leftrightarrow determiner

\neg (contracted_preposition \leftrightarrow preposition)

graduating_particle \rightarrow adjective

adjective \vee noun

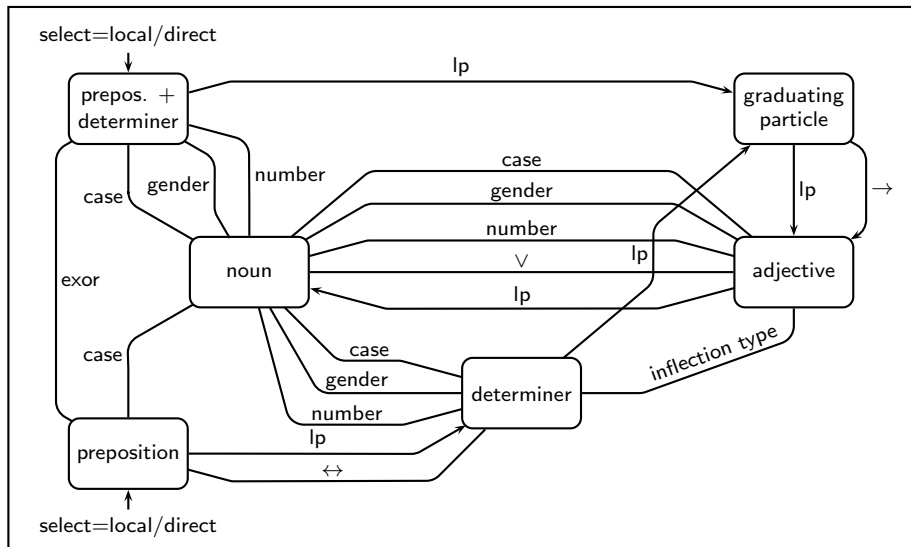
*auf Tisch

*in im Bett

*das sehr Auto

*wegen der

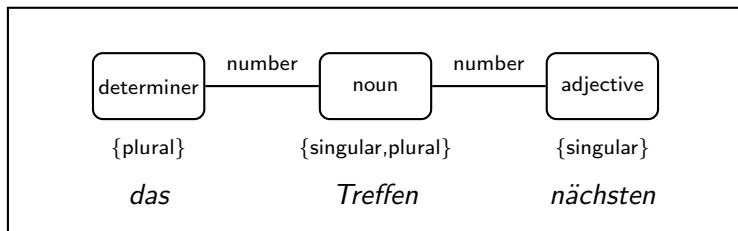
German Prepositional Phrase



- Constraint Satisfaction fails in case of ill-formed input
- By *retracting* constraints the global consequences of error hypotheses can be investigated
- Searching for *minimal* error hypotheses ...
- ... by successively increasing the number of retracted constraints

Diagnosis as Constraint Propagation

- Highly precise error explanations can be derived
- Different *views* on the error are supported
 - ▶ rule violations
 - ▶ missing lexical knowledge
- alternative error interpretations can be found

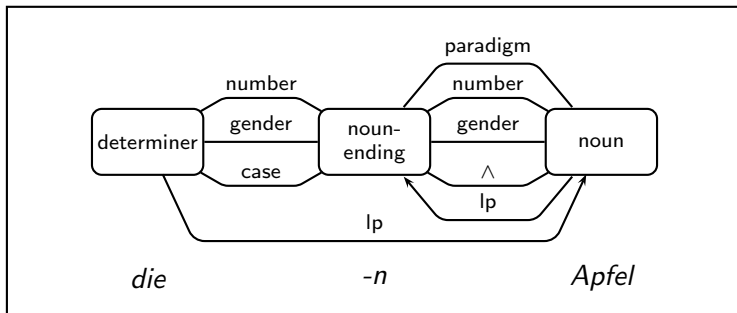


selecting an appropriate one according to the communicative context

- different feedback levels can be supported
 - ▶ error detection
 - ▶ error localization
 - ▶ error explanation
 - ▶ correction proposal

Diagnosis as Constraint Propagation

- Diagnosis of word formation errors



- error explanations for non-words become available

die Schachtel

der Apfel

die Schachteln

*die Apfeln

→ Apfel is masculine
not feminine

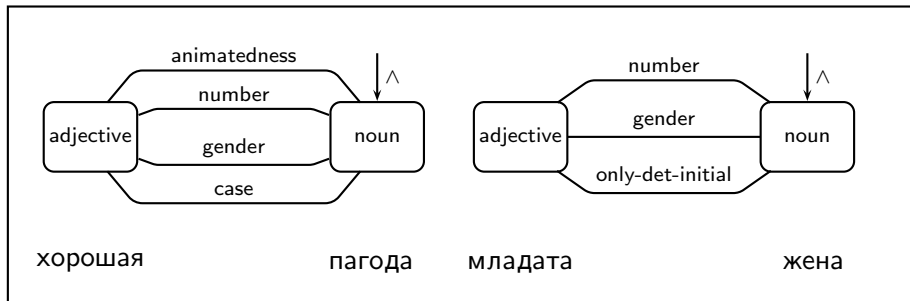
mit den Schachteln

*mit den Apfeln

→ the plural of Apfel
requires umlaut

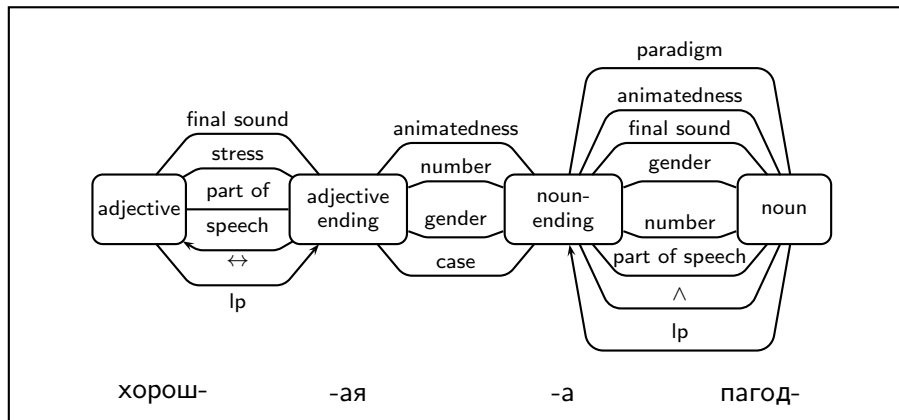
Diagnosis as Constraint Propagation

- Diagnosis in morphologically rich languages with full forms



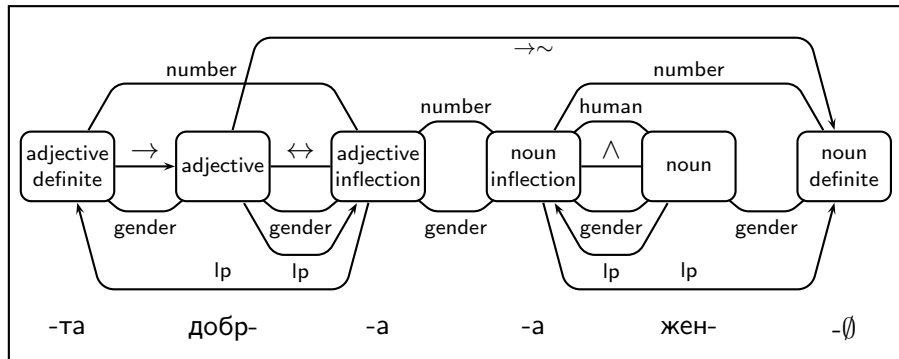
Diagnosis as Constraint Propagation

Morph-based diagnosis in Russian



Diagnosis as Constraint Propagation

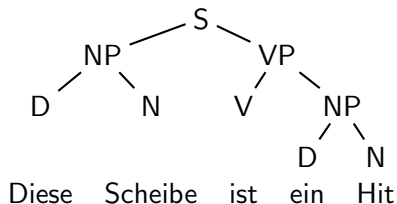
Morph-based diagnosis in Bulgarian



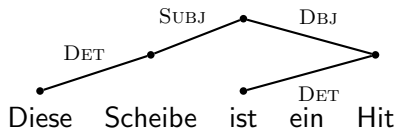
Structural Interpretation

- Parsing a natural language utterance means solving two tasks
 - ▶ finding structural descriptions
 - ▶ selecting the most plausible one
- Heuristic search in a large search space
- two different kinds of structural descriptions

phrase structure trees



dependency trees



Parsing as Constraint Satisfaction

- Labeled word-to-word are dependencies licensed by constraints
- Word forms correspond to the variables of a constraint satisfaction problem:
 - ▶ find the "correct" lexical reading
 - ▶ find the "correct" attachment point
 - ▶ find the "correct" label
- Parsing as structural disambiguation:
find a variable assignment which satisfies all constraints

Hypothesis space

root/nil	root/nil	root/nil	root/nil	root/nil
det/2	det/1	det/1	det/1	det/1
det/3	det/3	det/2	det/2	det/2
det/4	det/4	det/4	det/3	det/3
det/5	det/5	det/5	det/5	det/4
subj/2	subj/1	subj/1	subj/1	subj/1
subj/3	subj/3	subj/2	subj/2	subj/2
subj/4	subj/4	subj/4	subj/3	subj/3
subj/5	subj/5	subj/5	subj/5	subj/4
dobj/2	dobj/1	dobj/1	dobj/1	dobj/1
dobj/3	dobj/3	dobj/2	dobj/2	dobj/2
dobj/4	dobj/4	dobj/4	dobj/3	dobj/3
dobj/5	dobj/5	dobj/5	dobj/5	dobj/4
<i>Diese</i>	<i>Scheibe</i>	<i>ist</i>	<i>ein</i>	<i>Hit</i>
1	2	3	4	5

Parsing as Constraint Satisfaction

- Constraints license meaningful linguistic structures
- Natural language regularities do not depend on word positions
→ Constraints have to hold between arbitrary variables

$\{X\}$: DetNom : Det : 0.0 :

$X \downarrow \text{cat} = \text{det} \rightarrow X \uparrow \text{cat} = \text{noun} \wedge X.\text{label} = \text{DET}$

$\{X\}$: SubjObj : Verb : 0.0 :

$X \downarrow \text{cat} = \text{noun}$

$\rightarrow X \uparrow \text{cat} = \text{vfin} \wedge X.\text{label} = \text{SUBJ} \vee X.\text{label} = \text{DOBJ}$

$\{X\}$: Root : Verb : 0.0 :

$X \downarrow \text{cat} = \text{vfin} \rightarrow X \uparrow \text{cat} = \text{nil}$

$\{X, Y\}$: Unique : General : 0.0 :

$X \uparrow \text{id} = Y \uparrow \text{id} \rightarrow X.\text{label} \neq Y.\text{label}$

$\{X, Y\}$: SubjAgr : Subj : 0.0 :

$X.\text{label} = \text{SUBJ} \wedge Y.\text{label} = \text{DET} \wedge X \downarrow \text{id} = Y \uparrow \text{id}$

$\rightarrow Y \uparrow \text{case} = Y \downarrow \text{case} = \text{nom}$

- Natural language grammar is not fully consistent
- Many conflicting requirements
 - ▶ e.g. minimizing distance: verb bracket vs. reference

Sie hört sich die Scheibe, die ein Hit ist, an.



- Natural language grammar is not fully consistent
- Many conflicting requirements
 - ▶ e.g. minimizing distance: verb bracket vs. reference

Sie hört sich die Scheibe, die ein Hit ist, an.



Sie hört sich die Scheibe an, die ein Hit ist.



Conflicts occur

- between levels of conceptualization
e.g. syntax, information structure and semantics
- between different processing components
e.g. tagger, chunker, PP-attacher
- between the model and the utterance
e.g. modelling errors, not well-formed input
- between the utterance and the background knowledge
e.g. misconceptions, lies
- across modalities
e.g. seeing vs. hearing

Goal: achieve robustness and develop diagnostic capabilities

Why should we care about conflicts?

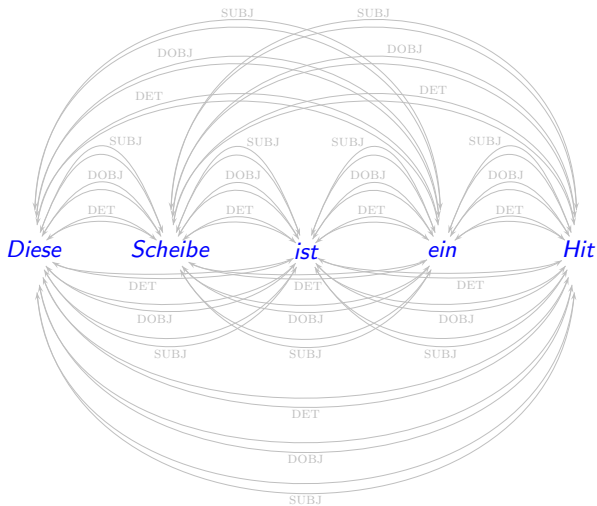
- they are pervasive
- they provide valuable information
 - ▶ for improving the system:
e.g. through manual grammar development or reinforcement learning
 - ▶ about the proficiency of the speaker/writer:
e.g. to derive remedial feedback
 - ▶ about the intentions of the speaker/writer:
e.g. attention focussing by means of topicalization
 - ▶ for guiding the parser

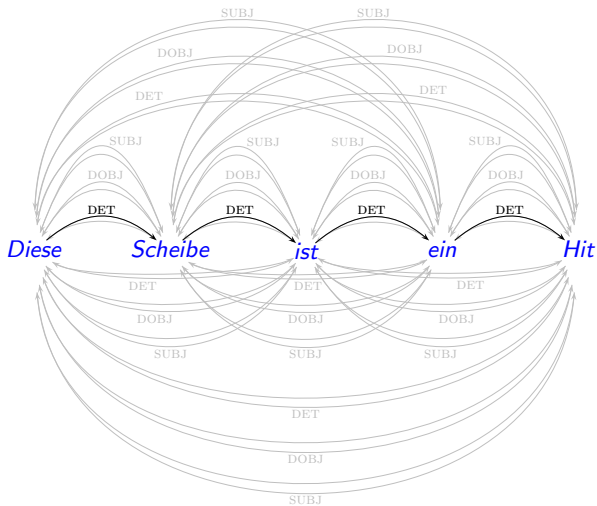
- conflict resolution requires weighted constraints
 - ▶ weights describe the importance of the constraint
 - ▶ how serious it is to violate the constraint
- differently strong constraints
 - ▶ hard constraints, must always be satisfied
 - ▶ strong constraints: agreement, word order, ...
 - ▶ weak constraints: preferences, defaults, ...

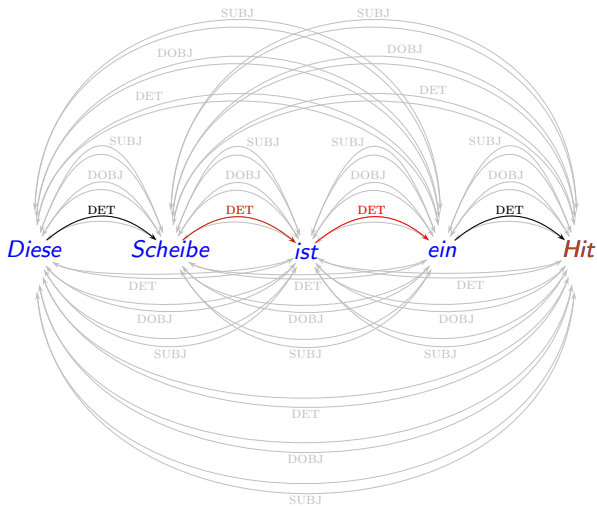
- weighted constraints are defeasible
- preferential reasoning can be applied
 - ▶ global optimization problem
 - ▶ based on local scores
 - ▶ scores are derived from constraint violations (penalties)

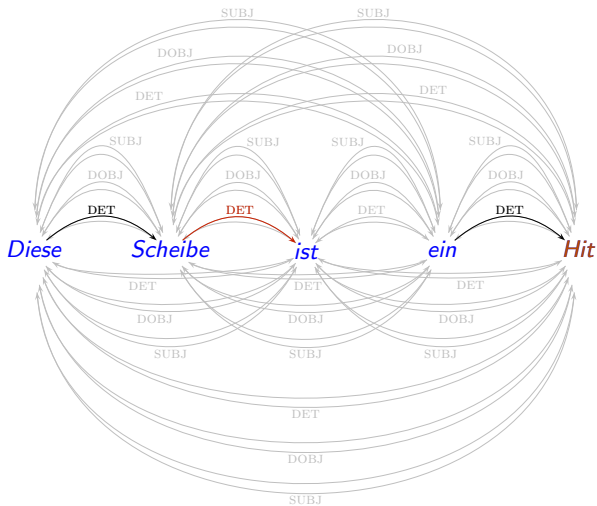
- Most constraints are local ones (unary, binary)
- Sometimes global requirements need to be checked
 - ▶ existence/non-existence requirements (e.g. valencies)
 - ▶ conditions in a complex verb group
- Local search supports the application of global constraints
 - ▶ always a complete value assignment (i.e. a dependency tree) is available
- Three kinds of global constraints
 - ▶ *has*: downwards tree traversal
 - ▶ *is*: upwards path traversal
 - ▶ recursive constraints: can call other constraints to be checked elsewhere in the tree

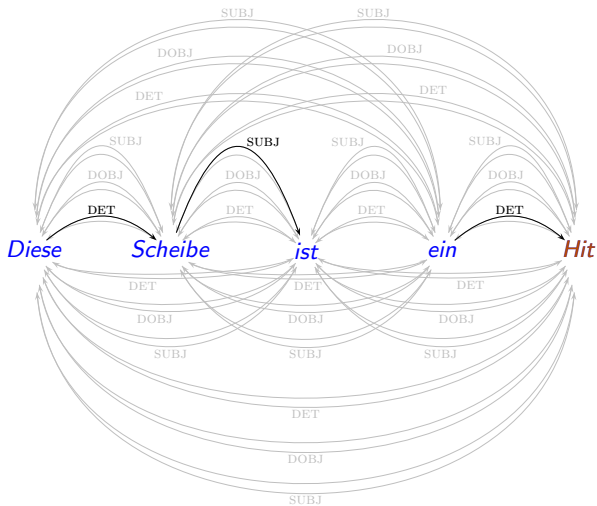
- Different solution procedures are available
 - ▶ pruning
 - ▶ systematic search
 - ▶ local search, guided local search (transformation-based)
- strong quality requirements
 - ▶ a single prespecified solution has to be found (gold standard)
 - ▶ sometimes the gold standard differs from the optimal solution
 - ▶ modelling errors vs. search errors
- The best method found so far:
 - ▶ local search with value exchange (frobbling)
 - ▶ gradient descent heuristics
 - ▶ with a tabu list
 - ▶ with limits (similar to branch and bound)
 - ▶ increasingly accepting degrading value selections to escape from local minima

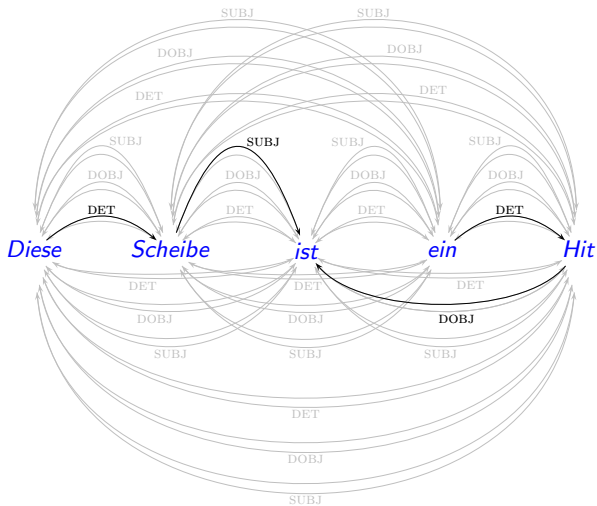












Non-local Transformations

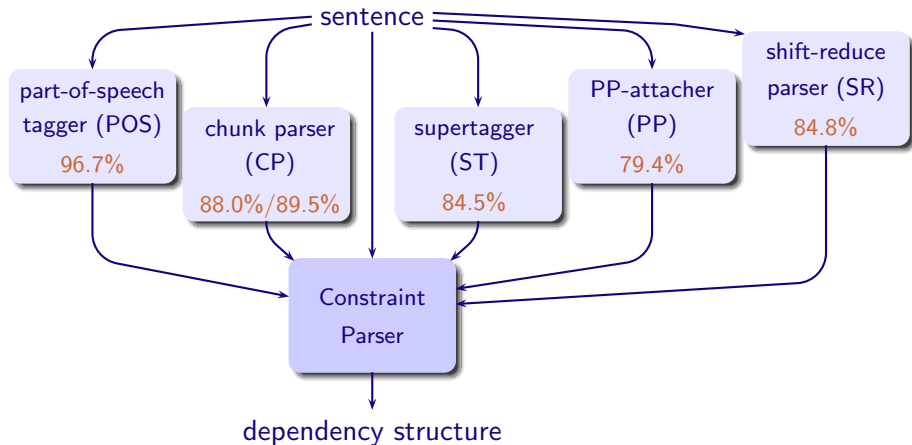
- usually local transformations result in unacceptable structures
- sequences of repair steps have to be considered.
- e.g. swapping SUBJ and DOBJ

a)	syntax	...	b)	syntax	...
diese ₁	det/2	...	diese ₁	det/2	...
scheibe ₂	dobj/3	...	scheibe ₂	subj/3	...
ist ₃	root/nil	...	ist ₃	root/nil	...
ein ₄	den/5	...	ein ₄	det/5	...
hit ₅	subj/3	...	hit ₅	dobj/5	...

The diagram illustrates a non-local transformation between two syntactic trees, labeled a) and b). In tree a), the word 'hit₅' is associated with the role 'subj/3'. In tree b), the word 'hit₅' is associated with the role 'dobj/5'. Red arrows show the movement of the 'dobj/3' role from tree a) to tree b) and the 'subj/3' role from tree b) to tree a), effectively swapping the subject and object roles for the word 'hit₅'.

- The bare constraint-based parser itself is weak
- But: constraints turned out to provide an ideal interface to external predictor components
- predictors might be inherently unreliable
→ can their information still be useful?
- using several predictors → consistency cannot be expected

Hybrid parsing

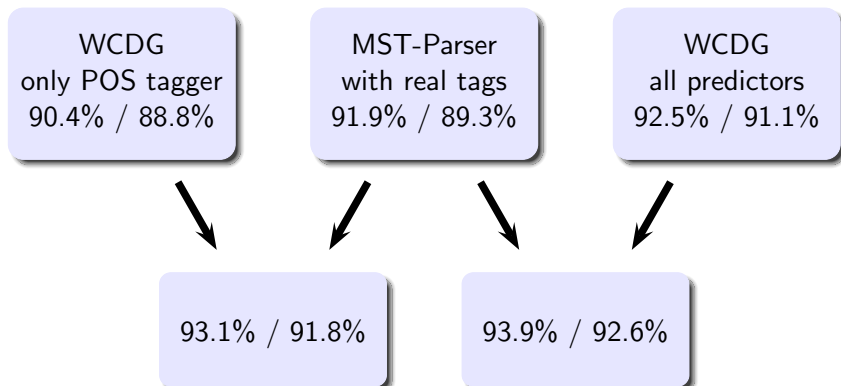


- results on a 1000 sentence newspaper testset (F_{OTH} 2006)

Predictors	accuracy	
	unlabelled	labelled
0: none	72.6%	68.3%
1: POS only	89.7%	87.9%
2: POS+CP	90.2%	88.4%
3: POS+PP	90.9%	89.1%
4: POS+ST	92.1%	90.7%
5: POS+SR	91.4%	90.0%
6: POS+PP+SR	91.6%	90.2%
7: POS+ST+SR	92.3%	90.9%
8: POS+ST+PP	92.1%	90.7%
9: all five	92.5%	91.1%

- net gain although the individual components are unreliable

- What happens if the predictor becomes superior?
(KHYMYLKO 2007)



Current research

- Incremental parsing
 - ▶ Language unfolds over time
 - ▶ Decisions about the optimal interpretation have to be taken in a timely manner
 - ▶ Local search has an ideal anytime behaviour: fully interruptable
- Parsing in a multimodal environment
 - ▶ Mapping visual stimuli onto linguistic constructions
 - ▶ Using language to guide the visual attention
- Using dynamic predictions
 - ▶ The world changes over time as the utterance unfolds
 - ▶ How does the behaviour of the parser depends on when an external information becomes available

Constraint satisfaction techniques ...

- simplify search problems
- provide diagnostic information
- can contribute attractive anytime properties

Weighted constraint satisfaction ...

- helps to solve hard optimization problems
- deals with conflicting regularities
- facilitates information fusion in hybrid architectures
- maintains the diagnostic abilities

Major challenge:

The search problem has to be recast in terms of a set of variables and their compatible value assignments.