Centering Theory and Algorithm

Definition | Characteristics | Architecture | Step-by-Step Procedure | Flow Graph | Detailed Example | Applications

Part	Title	Covers
1	Centering Theory Basics	Definition, constraints, rules, transitions, architecture, procedure, example, applications
2	Forward-looking (Cf) and Backward-looking (Cb) Centers	Definition, ranking, computation procedure, flow graph, detailed example, applications
3	The Centering Algorithm (BFP)	Procedure, architecture, full flow graph, 5-utterance worked example, Rule 1, variants, applications
4	Summary and exam questions	Comparison table, revision points, likely questions

Background: Read *'Ravi bought a dictionary. He had visited the shop many times. The shopkeeper knew him well.'* The reader's attention stays on **Ravi** across the three sentences, which is why 'He' and 'him' are easy to resolve. **Centering Theory** (Grosz, Joshi and Weinstein) models exactly this: *what the discourse is 'centrally about' at each point*, how that choice links consecutive utterances, and why some sequences feel coherent while others feel jumpy.

Part 1: Centering Theory Basics

1.1 Definition

Centering Theory (Grosz, Joshi and Weinstein, 1983; 1995) is a theory of **local discourse coherence**. It models the **attentional state** of the discourse participants within a discourse segment: it relates (i) the *focus of attention*, (ii) the *choice of referring expressions* (pronoun vs full noun phrase) and (iii) the *perceived coherence* of adjacent utterances. Each utterance has a set of **centers**, which are *entities* (not words) that link the utterance to the rest of the discourse segment.

Term	Meaning
Utterance (Un)	The unit analysed: usually a sentence or a finite clause
Discourse segment (DS)	A stretch of utterances sharing a purpose / topic; centering applies inside one segment
Center	A discourse entity (person, thing, concept) that links utterances; the same entity may be realised by a name, noun phrase or pronoun
Realization	An utterance *realizes* an entity if an NP in it denotes that entity (directly, or by bridging in extended versions)
Cf(Un)	Forward-looking centers: ranked list of entities realised in Un
Cb(Un)	Backward-looking center: the entity Un is 'about', linking back to Un-1
Cp(Un)	Preferred center: highest-ranked element of Cf(Un); predicts the next Cb

Place in Grosz and Sidner's discourse model

Component	Concerns	Level
Linguistic structure	Segmentation of the discourse into segments	Global
Intentional structure	Purposes of segments (DSPs) and their relations	Global
Attentional state (global focus stack)	Which segments / entities are in focus overall	Global
Centering	**Which entity is in focus from utterance to utterance inside a segment**	**Local**

1.2 Characteristics

- **Local, entity-based model of coherence:** works between adjacent utterances within one segment.
- **Attention is limited:** each utterance has at most **one** backward-looking center, the most prominent link to what came before.
- **Salience ranking:** entities are ranked, in English mainly by grammatical role (subject > object > others); the ranking is language specific.
- **Reference form matters:** highly prominent entities tend to be pronominalised (Rule 1).
- **Coherence is graded:** some transitions are cheaper for the hearer (less inference) than others.
- **Cognitively motivated:** aims to predict inference load; tested in reading-time and pronoun interpretation experiments.
- **Parameterised:** definitions of utterance, realization, ranking and previous-utterance can be set differently (Poesio et al. 2004).

The three constraints

No.	Constraint	Meaning
C1	Unique Cb	There is at most one Cb(Un) (it may be undefined)
C2	Realization	Every element of Cf(Un) must be realised in Un
C3	Cb choice	Cb(Un) is the **highest-ranked element of Cf(Un-1) that is realised in Un**

The two rules

Rule	Statement
Rule 1 (pronoun rule)	If *any* element of Cf(Un) is realised by a pronoun in Un, then Cb(Un) must also be realised by a pronoun.
Rule 2 (transition preference)	Transition states are ordered: **Continue > Retain > Smooth-shift > Rough-shift**. Sequences of continuations are preferred to sequences of retains, which are preferred to sequences of shifts.

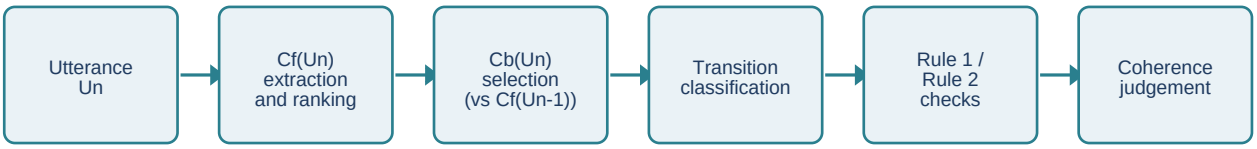
The four transitions

Condition	Cb(Un) = Cp(Un)	Cb(Un) != Cp(Un)
Cb(Un) = Cb(Un-1) (or Cb(Un-1) undefined)	**Continue** (keeps topic and plans to keep it)	**Retain** (keeps topic but Cp signals a coming shift)
Cb(Un) != Cb(Un-1)	**Smooth-shift** (shifts to the new, expected topic)	**Rough-shift** (abrupt shift, new topic not prominent)

1.3 Architecture

Conceptual architecture: an incremental model that carries a small attentional state from utterance to utterance.

Centering model



Component	Function
Input	A sequence of utterances U1...Un belonging to one discourse segment (with entities and coreference already identified or to be resolved)
State kept	Only Cf(Un-1) (ranked list) and Cb(Un-1): a very small, local memory
Ranking function	Language-specific salience order of entities (grammatical role, topic marker, information status)
Cb selector	Constraint C3: first element of Cf(Un-1) that occurs in Un
Transition classifier	Compares Cb(Un), Cb(Un-1), Cp(Un) to give Continue / Retain / Smooth-shift / Rough-shift
Rule checkers	Rule 1 (pronoun) and Rule 2 (preference order)
Output	Centers and transition sequence per utterance; coherence score or preferred pronoun interpretation

1.4 Step-by-step procedure (analysing a discourse with centering)

1. **Segment** the text into utterances U1, U2, ... Un (sentences or finite clauses) inside one discourse segment.
2. **Identify entities** in each utterance: names, noun phrases, and pronouns (resolve pronouns first or consider each possibility).
3. **Rank the entities** of each utterance to obtain Cf(Un) using the language's ranking (English: subject > object > object2 > others).
4. **Find Cp(Un)** = highest-ranked element of Cf(Un).
5. **Find Cb(Un)** = highest-ranked element of Cf(Un-1) that is realised in Un (Cb(U1) is undefined).
6. **Check Rule 1:** if any entity in Un is pronominalised, the Cb must be pronominalised.
7. **Classify the transition** from Un-1 to Un using the table above.
8. **Judge coherence** using Rule 2: more Continues give higher local coherence; many Rough-shifts or undefined Cbs signal weak local coherence.

1.5 Example (short)

```
U1: Meera wrote a letter. U2: She posted it. U3: The letter reached Anil the next day.
```

Utt.	Cf (ranked)	Cp	Cb	Transition	Explanation
U1	Meera, letter	Meera	undefined	-	First utterance

Utt.	Cf (ranked)	Cp	Cb	Transition	Explanation
U2	Meera (She), letter (it)	Meera	Meera	Continue	Cb = Meera (top of Cf(U1)); Cb = Cp; Rule 1 ok (She is a pronoun)
U3	letter, Anil, day	letter	letter	Smooth-shift	Meera not realised, so Cb = letter (next in Cf(U2)); Cb != Cb(U2) and Cb = Cp

The discourse moves smoothly from talking about Meera to talking about the letter; it stays coherent because the new Cb was already in the previous utterance and is also the most prominent entity of the new one.

1.6 Applications

- **Pronoun resolution:** prefer interpretations that give more coherent transitions (BFP algorithm, Part 3).
- **Coherence evaluation:** score essays, summaries and generated text by transition patterns (e.g., Miltsakaki and Kukich 2004 for essay scoring).
- **Natural language generation:** decide when to use a pronoun, how to order sentences and which entity to put in subject position.
- **Sentence ordering / text planning:** choose the order of facts that minimises shifts (Karamanis et al.).
- **Machine translation:** pronoun and zero-pronoun translation, especially Japanese to English where subjects are dropped.
- **Dialogue systems:** track the conversational focus to interpret 'it' and 'that'.
- **Psycholinguistics:** model reader's inference load and reading time for pronouns.

Part 2: Forward-looking and Backward-looking Centers

2.1 Definition

Forward-looking centers, Cf(Un): the set of discourse entities realised in utterance Un, *partially ordered by salience* (in English by grammatical role). They are 'forward-looking' because they are the candidates for linking the next utterance.

Backward-looking center, Cb(Un): the single entity that Un is 'about'; formally the highest-ranked element of Cf(Un-1) that is realised in Un. It links Un backward to Un-1.

Preferred center, Cp(Un): the highest-ranked element of Cf(Un); the best prediction of what Cb(Un+1) will be.

2.2 Characteristics

Property	Forward-looking centers (Cf)	Backward-looking center (Cb)
Cardinality	Many (list of all realised entities)	At most one (or undefined)
Direction	Looks forward: candidates for the next link	Looks backward: link to the previous utterance
Defined from	Utterance Un alone (plus ranking)	Both Un and Cf(Un-1)
Ordered?	Yes, ranked by salience	No ranking: it is a single choice
Realised by	Any NP: name, definite / indefinite NP, pronoun	Often a pronoun (Rule 1); may be any NP
Role	Salience information carried forward	Topic-like entity of the utterance; not necessarily the subject
First utterance	Defined	Undefined
Can be empty?	No (some entity is realised)	Yes: if no entity of Cf(Un-1) occurs in Un (no Cb), usually a topic change

Ranking schemes for Cf (language specific)

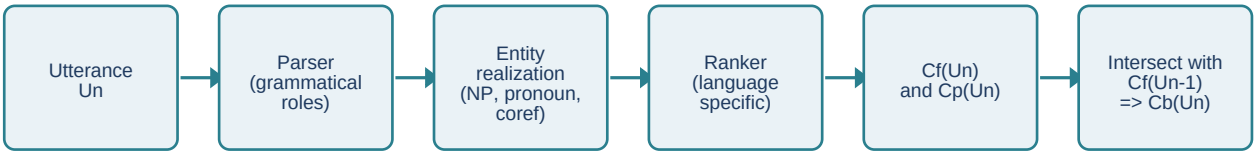
Language	Ranking of Cf (highest first)	Source
English	Subject > object > object2 (indirect / second object) > other complements and adjuncts	Brennan, Friedman and Pollard 1987
Japanese	Topic (wa) > empathy > subject (ga) > object2 > object > others	Kameyama 1986; Walker, Iida and Cote 1994
German and other free-word-order languages	Information-structure based: entities already known (hearer-old) ranked above new ones, then by position in the text	Strube and Hahn 1999

- **Cb is not the same as the subject:** in 'Ravi gave Meena a book. She thanked him.' the subject of U2 is Meena, but Cb(U2) = Ravi (highest-ranked entity of Cf(U1) realised in U2).
- **Cp is a prediction:** if Cp(Un) is later chosen as Cb(Un+1) the discourse is behaving as expected (Smooth-shift or Continue).
- **Locality:** Cb looks only at the immediately preceding utterance. An entity absent in Un-1 cannot be Cb(Un), even if it was prominent earlier.

- **Possessors and complements:** in extended versions, 'her sister' realises both 'sister' and the possessor 'her'.

2.3 Architecture of the Cf / Cb computation module

Cf / Cb computation

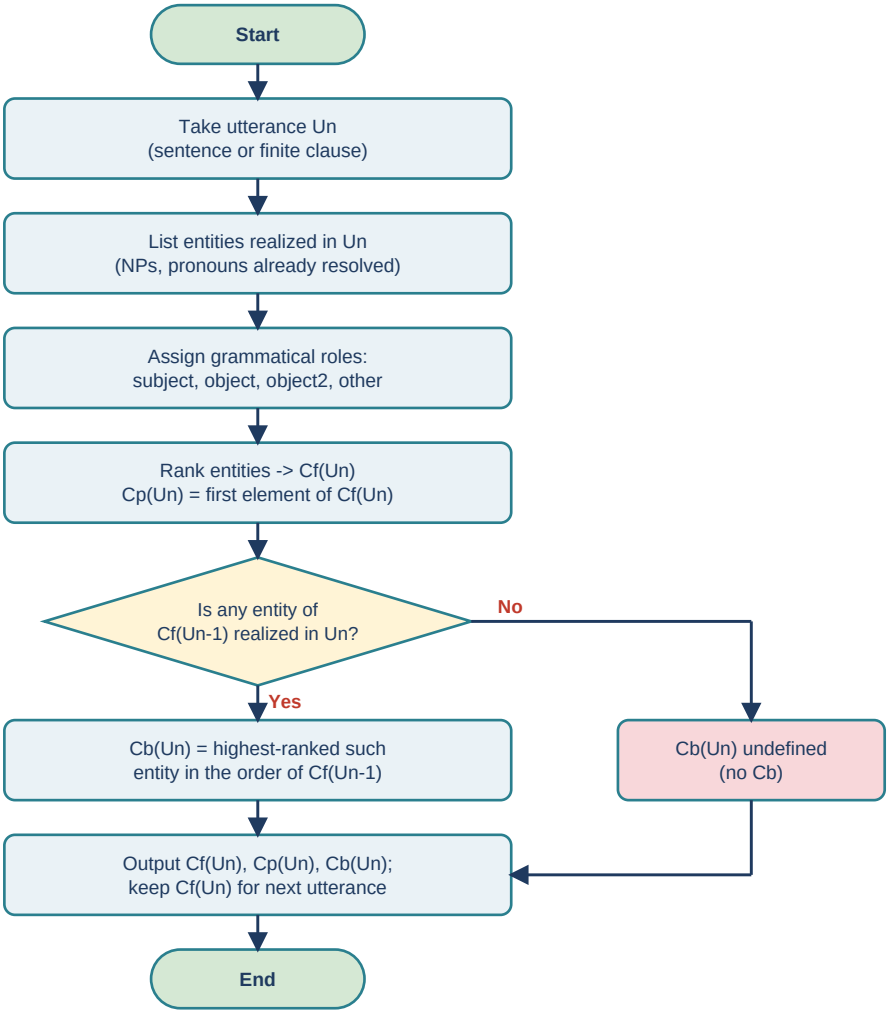


Module	Function
Parser	Gives grammatical relations (subject, object, indirect object, adjunct) or topic markers (Japanese wa)
Entity realization	Maps NPs and pronouns to discourse entities; needs coreference for pronouns and repeated names or definite NPs
Ranker	Orders entities by the language's salience hierarchy; ties broken by textual order
Cf store	Keeps Cf(Un-1) (only one previous utterance) and, in some variants, a short history
Cb selector	Scans Cf(Un-1) from the top; first entity also in Un becomes Cb(Un)

2.4 Step-by-step procedure to compute Cf, Cp and Cb

1. **Take the utterance Un** and parse it to identify its NPs and pronouns.
2. **List the entities realised** (resolve pronouns to entities first).
3. **Assign a role to each entity:** subject, object, second object, other.
4. **Sort by the ranking**: subject first, then object, second object, then others (left to right inside a class). The sorted list is **Cf(Un)**.
5. **Cp(Un)** = first entity of Cf(Un).
6. **Go through Cf(Un-1) from the top** and take the first entity that is also realised in Un: that is **Cb(Un)**. If there is none, Cb(Un) is undefined.
7. **Store Cf(Un)** and Cb(Un) for the next utterance.

2.5 Flow graph: computing Cf, Cp and Cb


Start
Take utterance Un
(sentence or finite clause)
List entities realized in Un
(NPs, pronouns already resolved)
Assign grammatical roles:
subject, object, object2, other
Rank entities -> Cf(Un)
Cp(Un) = first element of Cf(Un)
Is any entity of
Cf(Un-1) realized in Un?
No
Yes
Cb(Un) = highest-ranked such
entity in the order of Cf(Un-1)
Cb(Un) undefined
(no Cb)
Output Cf(Un), Cp(Un), Cb(Un);
keep Cf(Un) for next utterance
End

2.6 Example in detail

```
U1: Ravi gave Meena a book in the library.
U2: She thanked him.
U3: She read it at home.
U4: The library closes at nine.
```

Step A - U1: build Cf(U1)

Entity	Realised by	Grammatical role	Rank
Ravi	Ravi	Subject	1
Meena	Meena	Object (indirect)	2
book	a book	Object2 (direct)	3
library	in the library	Other (adjunct)	4

Cf(U1) = [Ravi, Meena, book, library]; Cp(U1) = Ravi; Cb(U1) = undefined (first utterance).

Step B - U2: 'She thanked him.' Resolve pronouns: She = Meena (fem), him = Ravi (masc). Roles: She = subject, him = object.

- Cf(U2) = [Meena, Ravi]; Cp(U2) = Meena.
- Cb(U2): scan Cf(U1) from the top: Ravi is realised in U2 (as 'him') -> **Cb(U2) = Ravi**.
- Notice Cb (Ravi) differs from the subject and from Cp (Meena).
- Rule 1: pronouns are used (She, him) and Cb is realised by a pronoun (him) -> satisfied.
- Transition: Cb(U1) undefined, Cb(U2) != Cp(U2) -> **Retain**. The speaker keeps Ravi as the link but Meena (Cp) is set up as next topic.

Step C - U3: 'She read it at home.' She = Meena, it = book, home = new entity.

- Cf(U3) = [Meena, book, home]; Cp(U3) = Meena.
- Cb(U3): scan Cf(U2) = [Meena, Ravi]: Meena is realised -> **Cb(U3) = Meena**. (Prediction confirmed: Cp(U2) = Meena became Cb(U3).)
- Rule 1: satisfied (She, it are pronouns; Cb = She).
- Transition: Cb(U3) = Meena != Cb(U2) = Ravi and Cb = Cp -> **Smooth-shift**.

Step D - U4: 'The library closes at nine.'

- Cf(U4) = [library]; Cp(U4) = library.
- Cb(U4): scan Cf(U3) = [Meena, book, home]: none is realised in U4. The library was in Cf(U1) but centering only looks at the previous utterance -> **Cb(U4) undefined**.
- Interpretation: a break in local coherence; the library is introduced as a new topic (or a new segment begins).

Utt.	Cf (ranked)	Cp	Cb	Transition
U1	Ravi, Meena, book, library	Ravi	undefined	-
U2	Meena, Ravi	Meena	Ravi	Retain
U3	Meena, book, home	Meena	Meena	Smooth-shift
U4	library	library	undefined	No Cb (break)

2.7 Applications of Cf and Cb

- **Pronoun resolution:** candidates are drawn from Cf(Un-1) in ranked order; the Cb-based preference selects among them.
- **Topic / focus tracking:** the Cb sequence gives the 'thread' of a text, usable for topic tracking and segment boundary detection (undefined Cb or repeated Rough-shifts suggest a boundary).
- **Generation:** put the entity to be continued into subject position (so it is top of Cf) and pronominalise the Cb.
- **Summarisation:** prefer sentence orders that keep the same Cb across sentences.
- **Entity-grid style coherence models:** Cf rank patterns are the basis of entity-based coherence features (Barzilay and Lapata).
- **Zero-pronoun resolution:** in Japanese, Hindi-style pro-drop, Cb / Cf ranking predicts the omitted subject.

Part 3: The Centering Algorithm (BFP)

3.1 Definition

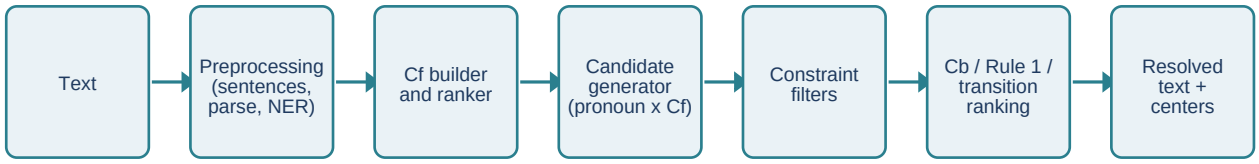
The **centering algorithm of Brennan, Friedman and Pollard (BFP, 1987)** is a pronoun resolution procedure built on centering theory. For each utterance it **generates all possible interpretations** of its pronouns, **filters** out those that violate constraints (agreement, binding, selectional restrictions, Rule 1), and **ranks** the rest by the **transition type** they produce (Continue > Retain > Smooth-shift > Rough-shift). The highest-ranked interpretation is chosen.

3.2 Characteristics

- **Incremental and local:** processes one utterance at a time with only Cf(Un-1) and Cb(Un-1) as memory.
- **Generate-filter-rank design:** exhaustive candidate generation, hard filters, then transition-based preference.
- **Handles pronouns only:** full NPs and names are assumed identified by other modules.
- **Uses linguistic knowledge:** parse trees (grammatical roles, binding), gender / number lexicons, selectional restrictions.
- **Deterministic and explainable:** every decision follows from a transition ranking, which makes errors easy to analyse.
- **Limited:** exponential growth of combinations with many pronouns; no semantic reasoning; sensitive to parsing errors and utterance definition; intra-sentential cases are hard.
- **Variants:** Left-Right Centering (Tetreault 2001) works incrementally left to right, usually more accurate and simpler; Strube's S-list approach ranks entities without Cb / transitions; Kameyama's treatment of complex sentences.

3.3 Architecture

BFP architecture



Module	Function
Preprocessing	Sentence / clause splitting, POS tagging, parsing for grammatical roles, NER, morphological features
Cf builder / ranker	Extracts entities of Un and orders them by role
Candidate generator	For each pronoun, takes entities of Cf(Un-1) (and of Un for intra-sentential) as candidates; builds the Cartesian product of choices (anchor combinations)
Constraint filters	Agreement (gender, number, animacy), syntactic disjoint reference (binding principles), selectional restrictions, Constraint C3 (Cb), Rule 1
Transition ranker	Classifies each surviving combination and selects the best (ties: first by Cf(Un-1) order)
State update	Commit the chosen interpretation; Cf(Un-1) := Cf(Un); Cb(Un-1) := Cb(Un)

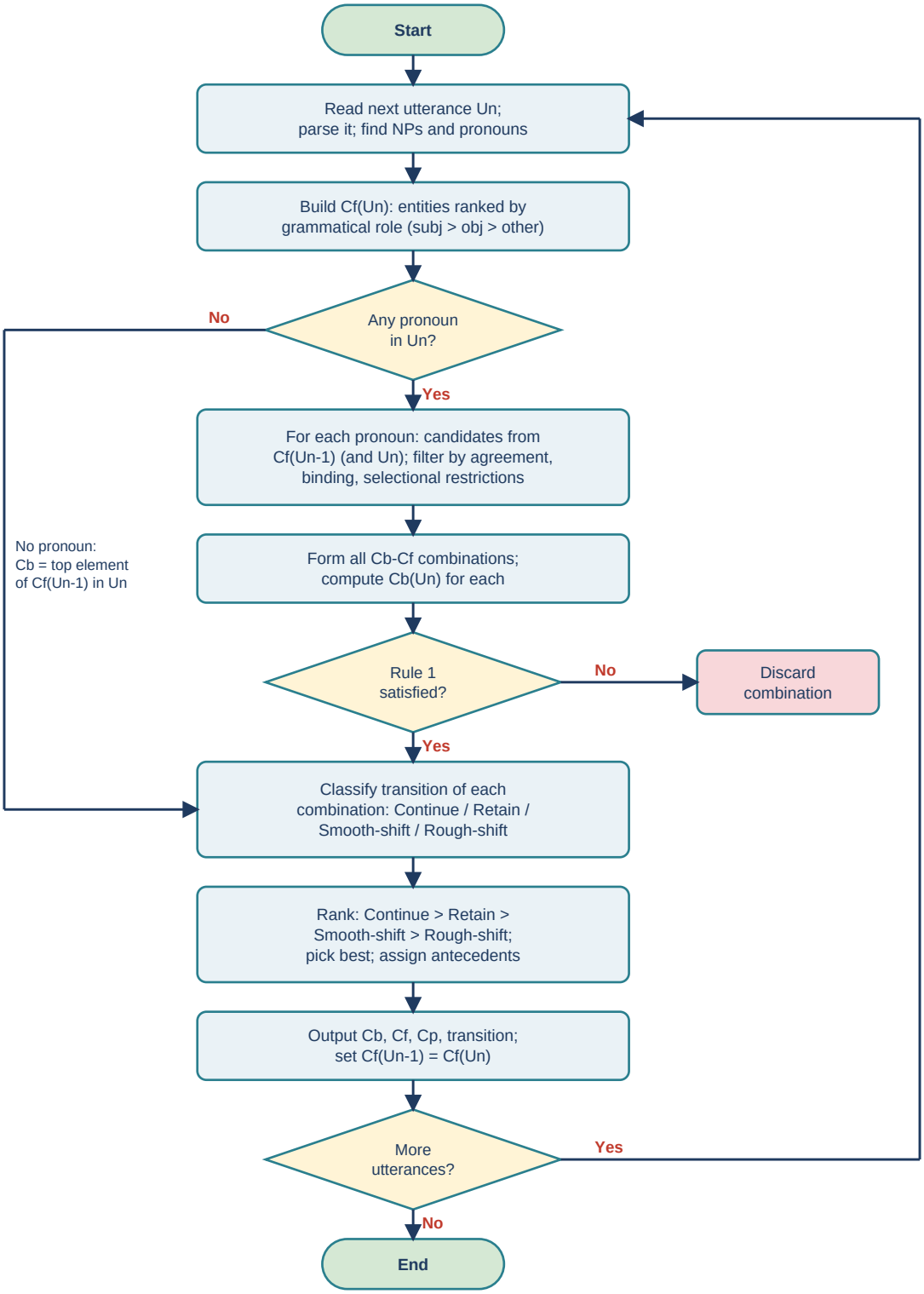
3.4 Step-by-step procedure

1. **Read utterance Un**; parse it; list NPs and pronouns; assign grammatical roles. For U1 just build Cf(U1) (Cb undefined).
2. **Generate candidates:** for each pronoun p in Un, candidate antecedents = entities in Cf(Un-1) (plus earlier entities in Un for intra-sentential reference).
3. **Filter by hard constraints:** remove candidates that clash with p in gender, number, animacy; that violate binding (e.g., 'him' cannot equal a clause-mate subject); or violate selectional restrictions.
4. **Form combinations:** take every consistent assignment of antecedents to all pronouns of Un. Each combination fixes the set of entities in Un.
5. **For each combination:** rank the entities to get Cf(Un); compute Cb(Un) (highest-ranked element of Cf(Un-1) realised in Un); compute Cp(Un).
6. **Apply Rule 1:** discard a combination if some element of Cf(Un) is a pronoun but Cb(Un) is not realised by a pronoun.
7. **Classify the transition** (Continue / Retain / Smooth-shift / Rough-shift) for each remaining combination.
8. **Choose** the combination with the most preferred transition (Continue > Retain > Smooth-shift > Rough-shift); break ties by Cf(Un-1) order.
9. **Commit** the resolution, set Cf(Un-1) := Cf(Un), Cb(Un-1) := Cb(Un), and go to the next utterance.

Pseudo-code

```
for each utterance Un:
  entities = parse(Un); roles = grammatical_roles(entities)
  for each pronoun p in Un:
    cand[p] = {e in Cf(Un-1) + earlier(Un) : agrees(p,e) and not
binding_violation(p,e)}
  combos = cartesian_product(cand[p] for all p) (drop combos that break disjoint
reference)
  for each combo c in combos:
    Cf_c = rank(entities under c); Cp_c = first(Cf_c)
    Cb_c = first e in Cf(Un-1) such that e in Cf_c (else undefined)
    if rule1_violated(c): discard c
    trans_c = classify(Cb_c, Cb(Un-1), Cp_c)
  best = combo with best trans (Continue > Retain > Smooth > Rough; ties by Cf order)
  commit(best); Cf(Un-1) = Cf_best; Cb(Un-1) = Cb_best
```

3.5 Flow graph of the centering algorithm



3.6 Detailed worked example

```
U1: Ravi bought a dictionary at the bookstore.
U2: He had visited the store many times.
U3: The shopkeeper knew him well.
U4: He gave him a discount.
U5: He was happy.
```

Assumptions: English ranking (subject > object > object2 > other); 'the store' is already resolved to 'bookstore' by a definite-NP resolver; all people are male, so agreement alone cannot separate them.

U1 - Ravi bought a dictionary at the bookstore.

- No pronouns. Entities and roles: Ravi (subject), dictionary (object), bookstore (other).
- **Cf(U1) = [Ravi, dictionary, bookstore]**; Cp(U1) = Ravi; Cb(U1) = undefined (first utterance). No transition.

U2 - He had visited the store many times.

Step	Action	Result
1. Pronouns	'He' (subject)	One pronoun
2. Candidates	From Cf(U1) = [Ravi, dictionary, bookstore]	Ravi, dictionary, bookstore
3. Agreement filter	He = masculine, animate	dictionary and bookstore removed -> only **Ravi**
4. Combinations	Only one	He = Ravi
5. Cf, Cp	Entities: Ravi (subject), bookstore (other)	Cf(U2) = [Ravi, bookstore]; Cp = Ravi
6. Cb	Scan Cf(U1): Ravi realised	**Cb(U2) = Ravi**
7. Rule 1	Pronoun He realises Cb	Satisfied
8. Transition	Cb(U1) undefined; Cb = Cp = Ravi	**Continue**

U3 - The shopkeeper knew him well.

Step	Action	Result
1. Pronouns	'him' (object); 'The shopkeeper' is a full NP, new entity	One pronoun
2. Candidates	From Cf(U2) = [Ravi, bookstore] plus the clause subject shopkeeper	Ravi, bookstore, shopkeeper
3. Filters	Agreement: bookstore removed. Binding (Principle B): 'him' cannot be the clause subject shopkeeper	only **Ravi**
4. Cf, Cp	shopkeeper (subject), Ravi (object)	Cf(U3) = [shopkeeper, Ravi]; Cp = shopkeeper
5. Cb	Scan Cf(U2) = [Ravi, bookstore]: Ravi realised, bookstore not	**Cb(U3) = Ravi**
6. Rule 1	Cb realised by pronoun 'him'	Satisfied
7. Transition	Cb(U3) = Cb(U2) = Ravi, but Cb != Cp (shopkeeper)	**Retain**

Prediction: Retain tells the reader that Cp(U3) = *shopkeeper* is a likely next Cb, i.e. a shift of attention to the shopkeeper is coming.

U4 - He gave him a discount. (the key ambiguity)

Step	Action	Result
1. Pronouns	'He' (subject), 'him' (indirect object); 'a discount' = object2	Two pronouns
2. Candidates	Cf(U3) = [shopkeeper, Ravi]: each pronoun can be either	2 x 2 = **4 combinations**
3. Agreement	All male, animate: nothing removed	4 remain
4. Binding	'He' and 'him' are co-arguments of 'gave', so they cannot corefer: remove (shopkeeper, shopkeeper) and (Ravi, Ravi)	**2 combinations**

	Combination A	Combination B
Interpretation	He = shopkeeper, him = Ravi	He = Ravi, him = shopkeeper
Cf(U4)	[shopkeeper, Ravi, discount]	[Ravi, shopkeeper, discount]
Cp(U4)	shopkeeper	Ravi
Cb(U4): first of Cf(U3) = [shopkeeper, Ravi] realised in U4	shopkeeper (realised as 'He')	shopkeeper (realised as 'him')
Rule 1	Cb is a pronoun: satisfied	Cb is a pronoun: satisfied
Compare with Cb(U3) = Ravi	shopkeeper != Ravi: a shift	shopkeeper != Ravi: a shift
Cb vs Cp	shopkeeper = shopkeeper: equal	shopkeeper != Ravi: not equal
Transition	**Smooth-shift**	**Rough-shift**

Decision: Smooth-shift > Rough-shift, so **He = shopkeeper, him = Ravi** is chosen. This agrees with common sense (the shopkeeper gives the customer a discount) and with the earlier Retain prediction. Commit: Cf(U4) = [shopkeeper, Ravi, discount], Cb(U4) = shopkeeper.

U5 - He was happy.

Step	Action	Result
1. Pronoun	'He' (subject)	One pronoun
2. Candidates	Cf(U4) = [shopkeeper, Ravi, discount]; agreement removes 'discount'	shopkeeper, Ravi: 2 combinations
3. Combination A: He = shopkeeper	Cf(U5) = [shopkeeper]; Cp = shopkeeper; Cb = shopkeeper (top of Cf(U4)); Cb(U4) = shopkeeper = Cb(U5); Cb = Cp	**Continue**
4. Combination B: He = Ravi	Cf(U5) = [Ravi]; Cp = Ravi; Cb = Ravi; Cb(U4) = shopkeeper != Cb(U5); Cb = Cp	**Smooth-shift**
5. Rank	Continue > Smooth-shift	He = **shopkeeper**

Final result

Utt.	Pronoun resolution	Cf (ranked)	Cp	Cb	Transition
U1	-	Ravi, dictionary, bookstore	Ravi	undefined	-
U2	He = Ravi	Ravi, bookstore	Ravi	Ravi	Continue

Utt.	Pronoun resolution	Cf (ranked)	Cp	Cb	Transition
U3	him = Ravi	shopkeeper, Ravi	shopkeeper	Ravi	Retain
U4	He = shopkeeper; him = Ravi	shopkeeper, Ravi, discount	shopkeeper	shopkeeper	Smooth-shift
U5	He = shopkeeper	shopkeeper	shopkeeper	shopkeeper	Continue

The Cb sequence (Ravi, Ravi, Ravi, shopkeeper, shopkeeper) shows attention staying on Ravi through U3 and then moving smoothly to the shopkeeper: a coherent discourse. Transition sequence: Continue, Retain, Smooth-shift, Continue, which is the typical natural pattern of a topic shift.

3.7 Rule 1 illustrated

Discourse	Analysis	Verdict
U1: Ravi gave Meena a pen. U2: Meena thanked **him**.	Cf(U1) = [Ravi, Meena, pen]. Cf(U2) = [Meena, Ravi]. Cb(U2) = Ravi (top of Cf(U1) realised) and it is a pronoun (him). Pronoun used, Cb is pronominalised.	Rule 1 satisfied: natural
U1: Ravi gave Meena a pen. U2: **She** thanked **Ravi**.	Cb(U2) = Ravi (realised by a name) while another entity (Meena) is pronominalised as 'She'.	Rule 1 violated: slightly awkward; 'She thanked him' is preferred

3.8 Limitations and variants

Issue	Explanation	Remedy / variant
Combinatorial explosion	n pronouns with k candidates give k to the power n combinations	Left-Right Centering (Tetreault 2001): process pronouns incrementally, first acceptable candidate
Pronouns only	Names, definite NPs, demonstratives are not resolved	Combine with coreference system
Utterance definition	Sentence vs clause changes Cf and Cb sets	Parameter studies (Poesio et al. 2004); Kameyama's complex sentence rules
Ranking is language dependent	English role order fails in free-word-order languages	Information-structure rankings (Strube and Hahn)
No semantics	Plausibility not used, so Winograd-style cases fail	Add selectional restrictions, commonsense models
Segment boundaries	Assumes the segment is known	Combine with topic segmentation

Reported accuracy: evaluations on newswire and fiction in the literature put the original BFP algorithm roughly in the 50-75 percent range for pronouns and Left-Right Centering and the S-list algorithm somewhat higher, around 75-80 percent; exact numbers depend on the corpus and the setup, so treat these as approximate and check the original papers.

3.9 Applications of the centering algorithm

- **Rule-based pronoun resolution** for English and other languages (Japanese, Italian, German, Hindi-style zero pronouns with adapted rankings).
- **Anaphora and coreference systems:** as a salience / focus component or feature (Cb match, transition type) in machine-learning resolvers.
- **Dialogue systems:** interpreting 'it', 'that', 'he' from the focus of the conversation.
- **Machine translation:** choosing target pronouns and recovering dropped subjects.
- **Natural language generation:** deciding when to pronominalise (pronominalise the Cb) and arranging sentences to keep Continue transitions.
- **Coherence assessment:** essay scoring, summary evaluation, sentence ordering based on transition counts.
- **Psycholinguistic modelling:** predicting which interpretation readers prefer for ambiguous pronouns.
- **Teaching and writing tools:** detect pronoun use that violates Rule 1 or too many Rough-shifts.

Part 4: Summary and Exam Questions

4.1 Quick revision

Item	Key points
Centering Theory	Local coherence model of attentional state in a discourse segment (Grosz, Joshi, Weinstein 1983; 1995)
Cf(Un)	All entities realised in Un, ranked (English: subject > object > object2 > others)
Cp(Un)	Top of Cf(Un); predicts Cb(Un+1)
Cb(Un)	Highest-ranked element of Cf(Un-1) realised in Un; undefined for U1 and when nothing is shared
Constraints	C1 at most one Cb; C2 all Cf elements realised in Un; C3 Cb = top Cf(Un-1) element realised in Un
Rule 1	If any element is pronominalised, the Cb must be pronominalised
Rule 2	Continue > Retain > Smooth-shift > Rough-shift
Transitions	Same Cb and Cb=Cp: Continue; same Cb and Cb!=Cp: Retain; new Cb and Cb=Cp: Smooth-shift; new Cb and Cb!=Cp: Rough-shift
BFP algorithm	Generate candidate combinations, filter by constraints and Rule 1, rank by transition, pick best
Variants	Left-Right Centering (Tetreault), S-list (Strube), parameterised centering (Poesio et al.)
Uses	Pronoun resolution, coherence scoring, generation, sentence ordering, MT

4.2 Likely exam questions

1. State and explain Centering Theory. What are its constraints and rules?
2. Define forward-looking center, backward-looking center and preferred center with examples. How is the Cf list ranked?
3. Explain the four centering transitions with a table and an example for each.
4. Describe the BFP centering algorithm with a flow graph and an example showing how an ambiguous pronoun is resolved.
5. Explain Rule 1 with an example. What happens when it is violated?
6. Compare centering with other salience approaches (Hobbs, Lappin and Leass, Left-Right Centering).
7. Discuss limitations of centering theory and how modern systems overcome them.
8. List the applications of centering theory in NLP.

Note: the definitions, constraints, rules and ranking schemes follow the standard presentation of Grosz, Joshi and Weinstein (1995) and Brennan, Friedman and Pollard (1987); all examples and worked sentences were written for these notes. The accuracy range quoted in 3.8 is approximate and from memory of the literature: check the original papers or your textbook before quoting figures in an exam.