# Syntactic mismatches in French Peripheral ellipsis corpus study and experiments

Anne Abeillé,Barbara Hemforth

Aoi Shiraïshi

University Paris Diderot

May 20th 2015

# PLAN

– 1 Peripheral ellipsis (PE)

   or Right-Node Raising

– 2 Det mismatch in French PE

– 3 Prep mismatch in French PE

– 4 Voice mismatch in French PE

– 5 Analyses of PE
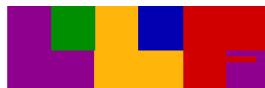
# 1. Peripheral ellipsis

Usually called Right node raising (RNR)

*John likes but Mary dislikes* *bananas.*

Can occur outside coordination or dialog:

*Anyone who meets really comes to like our sales people (*Williams 1990)

Can apply to non maximal constituents:

*It was a sweet and an intelligent dog* (SWB corpus)

Can apply to wordparts (Chaves 2008):

*These events took place in pre- or in post-war Germany* *?*

# 1.1. Syntactic mismatches in ellipsis

**Syntactic mismatches**:

 when the source and the target have different syntactic categories, features, dependents…

⇒Semantic reconstruction or at LF

**Well-known for VP Ellipsis**:

=> Argument for a null complement analysis of VP ellipsis (Hardt 1993, Ginzburg & Sag 2000…)

# VP Ellipsis and mismatch

<u>Tense mismatch</u>

*I have looked into this problem and you should.* <span style="color:red">*(look into that problem*</span>*)*

<u>Voice mismatch</u>

*This problem was to have been looked into, but obviously nobody did.* (<span style="color:red">look at it</span>) (Hardt 1993, Kehler 2000, Kertz 2014…)

<u>Category mismatch</u> (Kehler 2000, Miller & Hemforth 2014)

*This letter deserves a response, but before you do…* (<span style="color:red">respond</span>….)

# Peripheral Ellipsis (cont)

Is usually assumed to impose strict identity conditions

*I know that Paul ~~is leaving~~ but I don't know whether his children <span style="color:red">are leaving</span>

*Paul saved ~~himself~~, but Mary didn't save <span style="color:red">herself</span>.

Mismatch said to require syncretic forms (Pullum & Zwicky 1986)

* I already have ~~clarified the situation~~ and you certainly will <span style="color:red">clarify the situation</span>

I already have ~~set the record straight~~ and you certainly will <span style="color:red">set the record straight</span>

Laboratoire de Linguistique Formelle

# 1.2 Corpus studies on PE

## Bilbîie 2013 Penn treebank (gapping and RNR)

RNR does occur outside coordination:

*Nearly half of them argue that Hong Kong's uneasy relationship with China will constrain – though not inhibit – long-term economic growth.* (wsj-30583)

RNR is mostly subclausal :

*This was a formal or an informal dinner party* ? (swbd-132959)

RNR often involves polarity or modality contrast:

*Did you or did you not say what I said you said*... ? (brwn-4498)

*Who is and who should be making the criminal law here* ? (wsj-6370)

# Previous corpus study on French PE

Abeillé and Mouret (2010):

- French treebank (Abeillé et al 2003)(1 M words newspaper): corpus syntactically annotated but not for ellipsis

-  Ester (Gravier et al. 2004)(a radio corpus tagged for POS)

=> manual search

– Results: mostly subclausal ellipsis, with contrast on tense, modality, polarity
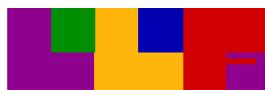
# Abeillé & Mouret 2010

## Polarity mismatch:

*Il y a des langues qui ont ~~une flexion casuelle~~, et des langues qui n'ont pas, de flexion casuelle* (C. Hagège) ('there are languages that have and languages that don't have case inflection')

## Preposition mismatch:

Une personne sur 3 est incapable ~~de mener une vie indépendante~~ ou a beaucoup de mal *à mener une vie indépendante* (France Inter, Ester) '1 person out of 3 is unable or has much trouble to lead an independant life'

*Ce parti ne parvient pas ~~à surmonter ses contradictions~~, voire ne souhaite pas, surmonter ses contradictions*. (Le Monde FTB)

(This party doesn't succeed, nor doesn't want to overcome its contradictions)

# 2.Det mismatch in French pE

No French corpora annotated for ellipsis

Our search (on the web):

Regular expressions with pronominal subjects and frequent verbs:

*qu'il/elle/on ait/fasse/prenne/mette ou qu'il/elle/on n'ait/ fasse/prenne/mette pas de*

*qu'ils/elles aient/fassent/prennent/mettent ou qu'ils/ elles n'aient/ne fassent/prennent/mettent pas de*

*s'ils/elles ont /font/prennent/mettent ou s'ils /elles n'ont/ font/prennent/mettent pas de*

# Determiner mismatch in PE

*de*: polarity variant of indefinites *un/du/des*

Analysed as marker by Abeillé et al. 2003

*une/de*

*Les textes actuels permettent de citer à l'audience une personne, qu'elle ait* ~~*une dernière adresse connue*~~ *ou qu'elle n'ait pas de dernière adresse connue*. (Avis Conseil d'état, Feb 13rd 2007)

(The current texts allow one to cite in court a person, whether she has or she doesn't have any known adress)

# Det mismatch

## *du/de*

Que la consommatrice cherche ou ne cherche pas un produit, qu'elle ait ~~du mal à le localiser~~ ou qu'elle n'ait pas de mal à le localiser (…) (Franck Cochoy, Les figures sociales du client, 2002)

*'…whether she has or she doesn't have any trouble to locate it…'*

## *des/de*

*C'est de la responsabilité de l'Eglise de venir en aide aux migrants et aux réfugiés qu'ils aient ~~des papiers~~ ou qu'ils n'aient pas de papiers* (Mgr Dognin, Tours, Aug 1st 2014)

'It is the Church's responsibility to help migrants and refugees whether they have or don't have papers'

* qu'ils aient de papiers
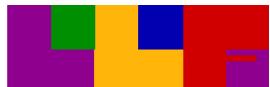
# Determiner mismatch in PE

with definite determiner :

Que je sache, même s'il rencontre ~~le pape François~~ ou s'il téléphone *au pape François*, il ne prend pas sa place.

(lepeupledelapaix.forumactif.com/t20280-medaillons-des-papes-il-en-rest...)

'…even if he meets or if he phones Pope Francis…'

with  bound possessive:

Je parle (…) de tous ceux qui se sont battu pour que je ~~puisse ouvrir ma gueule~~ et que tu puisses *ouvrir ta gueule* en toutes liberté dans un pays libre (July 16 2009
http://mouvement-ultra.forumactif.fr/)

'…so that I and taht  you can open your mouth…'
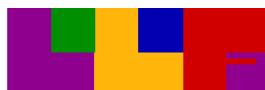
# 3. Preposition mismatch in PE

*à/de* + infinitive

Manual search on the internet with frequent verbs*: tenter de, réussir à, essayer de, parvenir à, être en train de*

*qui/s'il/si elle/on tente ou qui/qui/s'il/si elle/on réussit à…*

*qui/s'il/si elle/on tente ou qui/qui/s'il/si elle/on réussit à…*

Sandra Gons fait ressurgir dans ses dessins les non-dits d'une société qui ne parvient pas ~~à réaliser son introspection~~ – ou ne veut pas - réaliser son introspection, (http://www.liberation.fr/monde/2014/02/02/)

'…a society which doesn't succeed, nor doesn't want to achieve its introspection…'

# Preposition mismatch

à/de + NP

Pour moi, un français qui va à Honarribia ou qui revient d'Hondarribia ne verra que des panneaux et des cartes avec "Hondarribia" (discussion wikipaedia 2007)

'For me, a French man who goes or who comes from Hondarribia will only see signs and maps with Hondarribia'

– Can be combined with det mismatch:

les brancardiers qui sillonnent les couloirs [..] avec toujours un sourire ou un mot rassurant pour un malade qui va au bloc ou qui revient du bloc.(blog 2015)

'...with always a smile or reassuring word for a patient who goes or who comes back from the operating room'

# Are these mismatches grammatical ?

- – There are RNR with mismatches in news sites, books and official notes.

- – Are these examples grammatical ?

- – We therefore have conducted an acceptability judgment test in order to test their acceptability.
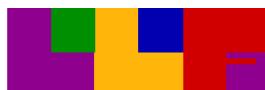
# Target items for deteminer mismatch

Experiment with sentences (from attested examples) in 3 conditions :

(a) Il y a des gens qui ont, et des gens qui n'ont pas, *de problème de poids*. ellipsis with mismatch

(b) Il y a des gens qui ont un problème de poids, et des gens qui n'en ont pas. No ellipsis

(c) Il y a des gens qui ont, et des gens qui n'ont pas, *un problème de poids*.      ellipsis without mismatch

# Target items for preposition mismatch

Experiment with sentences (from attested examples) in 3 conditions :

(a) Que faire si votre conjoint ne parvient plus, ou s'il ne veut plus, *vous parler* ? <span style="color:blue">ellipsis with mismatch</span>

(b)  Que faire si votre conjoint ne parvient plus à vous parler, ou s'il ne veut plus ? <span style="color:blue">No ellipsis</span>

(c) Que faire si votre conjoint ne peut plus, ou s'il ne veut plus, *vous parler* ? <span style="color:blue">ellipsis without mismatch</span>
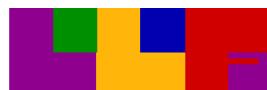
# Control items

1. *Jean a le courage de ses opinions*    grammatical
2. *Jean a courage de ses opinions*    ungrammatical
3. *Jean a de courage de ses opinions*    ungrammatical

# Acceptability judgment test

– 48 test items (preposition and determiner mismatch, RNR and lexical coordination) and 15 control items

– 40 subjects (native) : rating (1-10) on the RISC platform

# Results

## Experiment 1: Results (Det & Prep)



**Mismatch**

**No ellipsis**

**No mismatch**

**Grammatical controls**

**Ungrammatical controls**
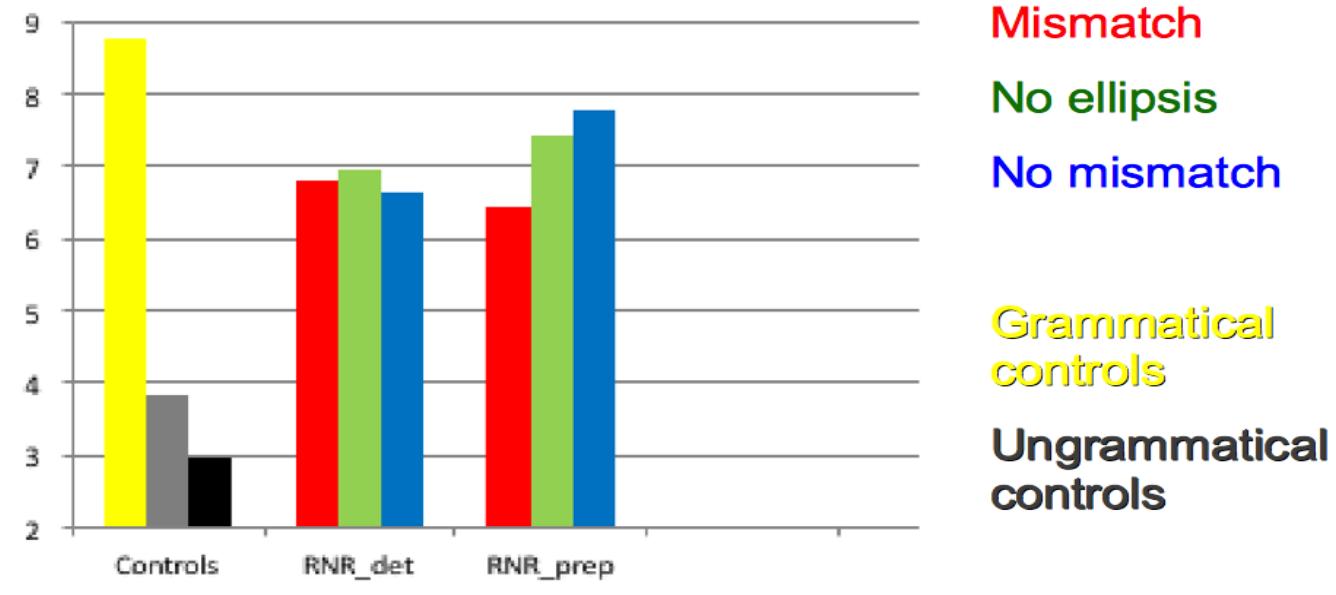
# Results: determiner mismatch
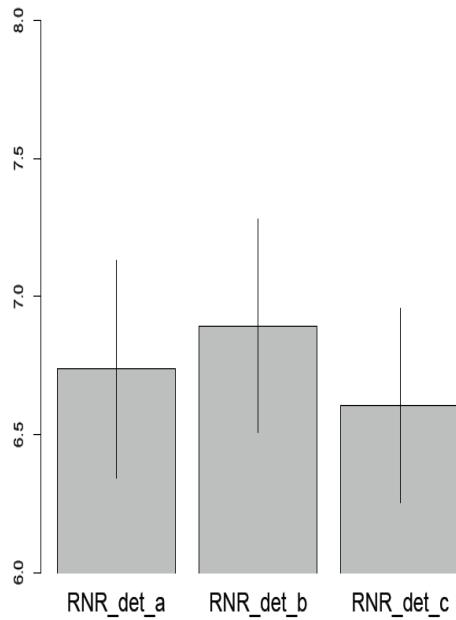


- RNR with det mismatch
  (a) rated at 6.779 is as
  acceptable as RNR
  without det mismatch
  (c) rated at 6.6624

- No significant difference
  between RNR with
  mismatch (a) and RNR
  without (c)

# Results: preposition mismatch
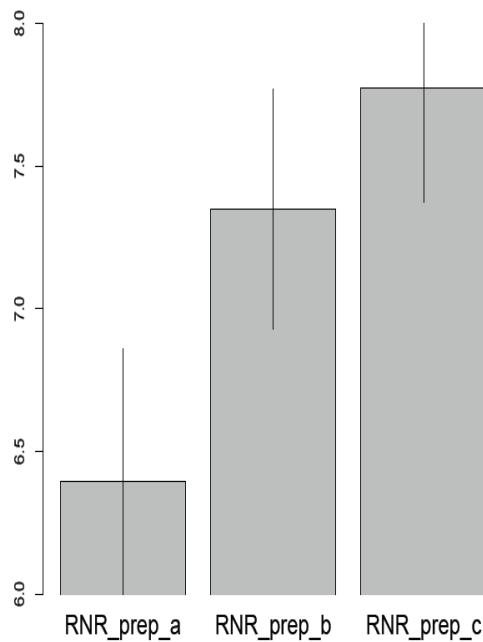


- RNR with prep mismatch (a) rated at 6.445 is less acceptable than RNR without prep mismatch (c) rated at 7.77.

- no significant difference between Coord without RNR (b) and RNR without prep mismatch (c)

# Results

-Ellipsis with determiner mismatch is more acceptable than without (surprising but mismatch examples come form corpora);

-Ellipsis with preposition mismatch is less acceptable than without, but much higher than ungrammatical controls;

In case of mismatch, the requirement of the second clause wins

⇒ difficult for raising or multidominance approaches

# 4. Voice mismatch in PE

Past and passive participles are syncretic forms

No French corpus annotated for ellipsis=> manual search

A lot of attested examples with tense or modality contrast:

[Fernand Deligny] Il utilise l'expression par le dessin, le jeu mimé, l'alphabet-geste, le récit improvisé collectivement, pour donner parole à ceux qu'on a ~~privés de dire~~ ou qui sont privés de dire. www.cemea.asso.fr/ avril 1997

('…those that one has or who are deprived of talking')

– Épousée , s, /. celle qu'on a ~~épousée~~, ou qui doit être bientôt épousée. 'the one that one has - or that must soon be married' Dictionnaire universel de la langue françoise, PCV Boiste 1803
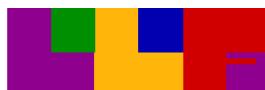
# Other examples with role contrast

Refl active-passive

(1) Ce pharmacien doit des explications à ceux qui se sont ou qui ont été <span style="color:red">mobilisés pour lui</span> (March 2013, www.ipreunion.com/)

'This pharmacist owes explanations to those who have themselves or who have been rallied for him'

Passive- refl active

(2) Mais il y a aussi, tous ceux qui ont été ou qui se sont <span style="color:red">exclus</span>, plus tard, dans leur vie d'adulte. (2008, www.ville-yzeure.com/)

'But there are all those who have been or have themselves excluded in their adult life'

Intrans active-passive

(3) dans n'importe quelle démocratie, le ministre de l'intérieur aurait, ou aurait été <span style="color:red">démissionné</span> (Le Figaro, sept 2014) 'in any democracy, the interior Minister would have or would have been resigned'

# Could that be VP ellipsis?

- Cataphoric VP ellipsis involve subordination

*If you can, you should leave now.*

- No VP ellipsis in French:

*\*Jean a démissionné mais Marie n'a pas.*

'Jean has resigned but Marie has not.'

*\*Certains ont été exclus mais d'autres n'ont pas été.*

'Some have been excluded but others have not been.'

No VP complement after French auxiliaries (Abeillé & Godard 1994)

# Semantic contrast in PE

French PE needs semantic contrast (on predicate, tense, mode, polarity…)

*Elle s'appelait, et elle s'appelle encore, <span style="color:red">Christine Sprenger</span> (*France Inter, Ester) 'she was called and she is still called CS'

Good mismatches involve tense, mode, or role contrast

Bad mismatches involve contrastive topics (Kertz 2014):

*# Paul a été, mais son frère a <span style="color:red">convaincu</span>.*

'Paul has but his brother has been convinced'

Or no contrast:

# ceux qui ont été ou qu'on a convaincus

(those that one has or that have been convinced)
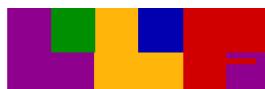
# Experiment1 (voice mismatch)

+contrast/-match

a. Il s'agit d'Eglises orientales qui se sont, ou qui ont été rattachées à Rome à un moment donné.

-contrast/-match

b. Il s'agit d'Eglises orientales qu'on a, ou qui ont été rattachées à Rome à un moment donné.

+contrast/+match

c. Il s'agit d'Eglises orientales qui se sont, ou qu'on a rattachés à Rome à un moment donné.

-contrast/+match

d. Il s'agit d'Eglises orientales qui étaient, ou qui ont été rattachées à Rome à un moment donné.

Acceptability rating (online) : 12 items; 65=> 46 subjects
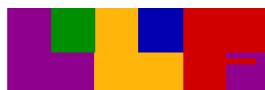
# Results
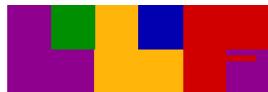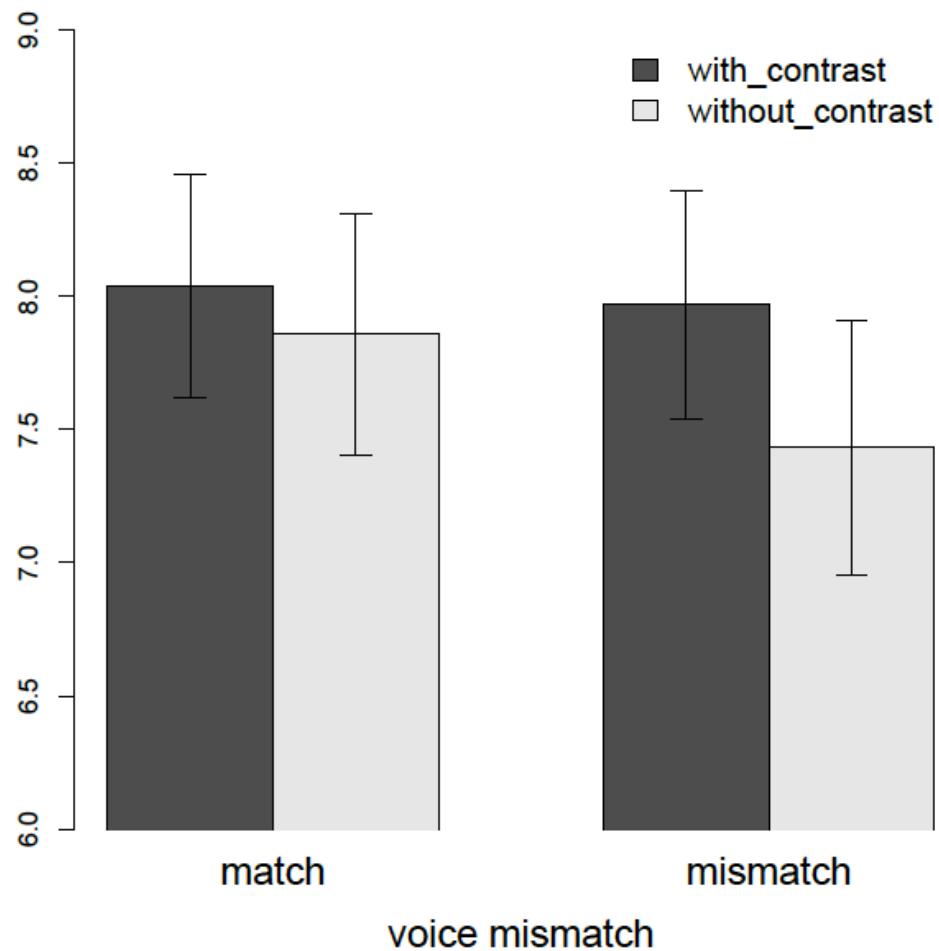
# Results

Only the semantic condition played a role:

Mean match a: 8,144927536        b:7,666666667

   mismatch c:  8,217391304       d:8,036231884

Mixed linear regression model

No significative effect of match*contrast

Model without interaction:

No significative effect of match : 0.6649

Significative effect of contrast: 0.0495*

With standardized values:

Contrast: t -2.215*   p 0.027*      match: 0.93

# Experiment2. Eye Tracking

extension of  material of experiment1

<u>+contrast/-match</u>

a. Ce pharmacien doit des explications à ceux qui se sont, ou qui ont été mobilisés pour sa défense.

<u>-contrast/-match</u>

b. Ce pharmacien doit des explications à ceux qu'on a, ou qui ont été mobilisés pour sa défense.

<u>+contrast/+match</u>

c. Ce pharmacien doit des explications à ceux qui se sont, ou qu'on a mobilisés pour sa défense.

<u>-contrast/+match</u>

d. Ce pharmacien doit des explications à ceux qui étaient, ou qui ont été mobilisés pour sa defense.
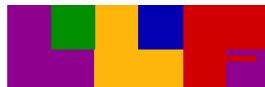
# Eye tracking experiment

## -20 items, 36 subjects

-thanks to Dario Paape and Celine Pozniak

Eliminate outliers (too long reading times or too short reading times)

## -Concentrate on Interest Area 4

Ce pharmacien doit des explications à ceux qui se sont, *ou qui ont été* mobilisés pour sa défense.

# Graph from course
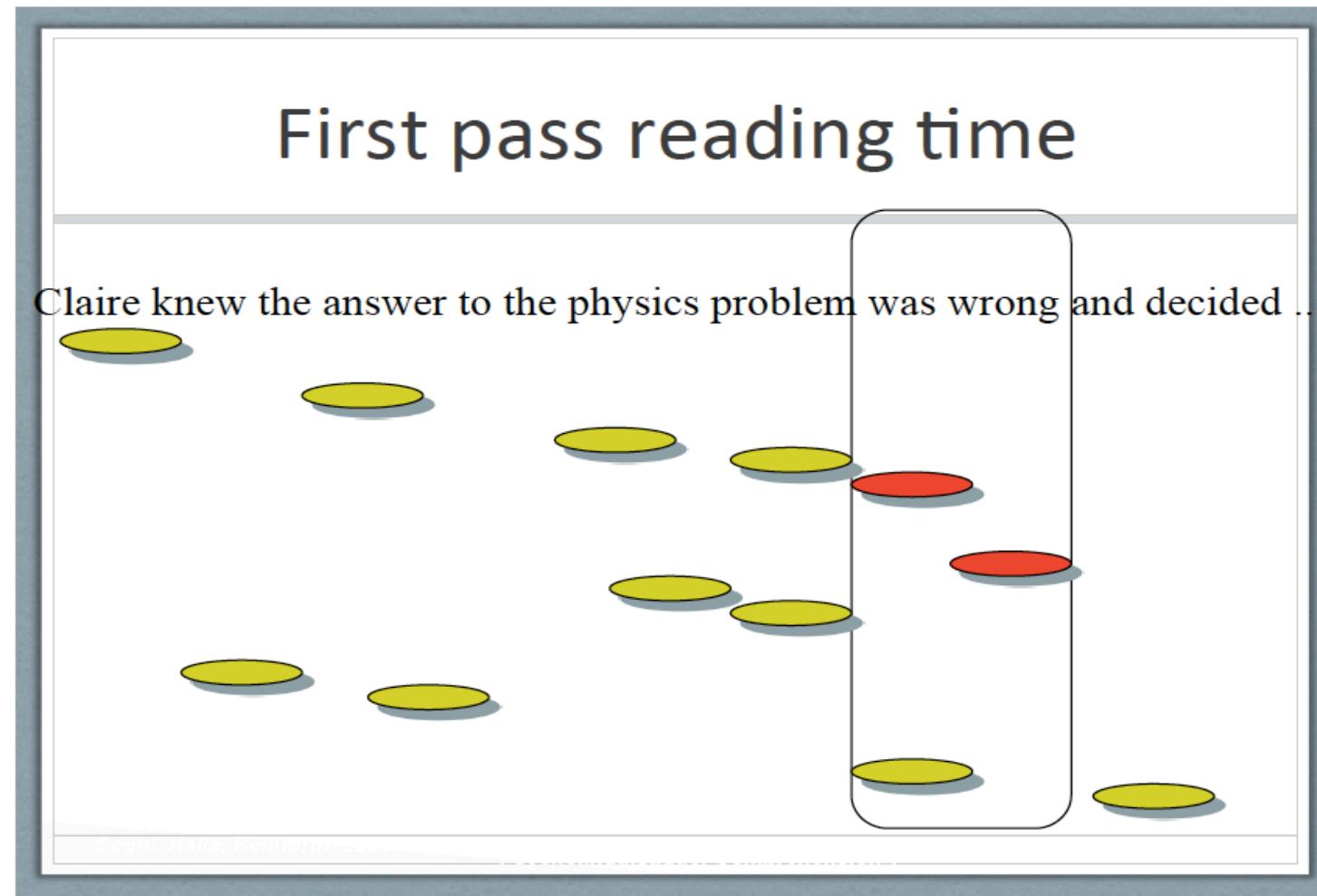
# First Pass Reading Times
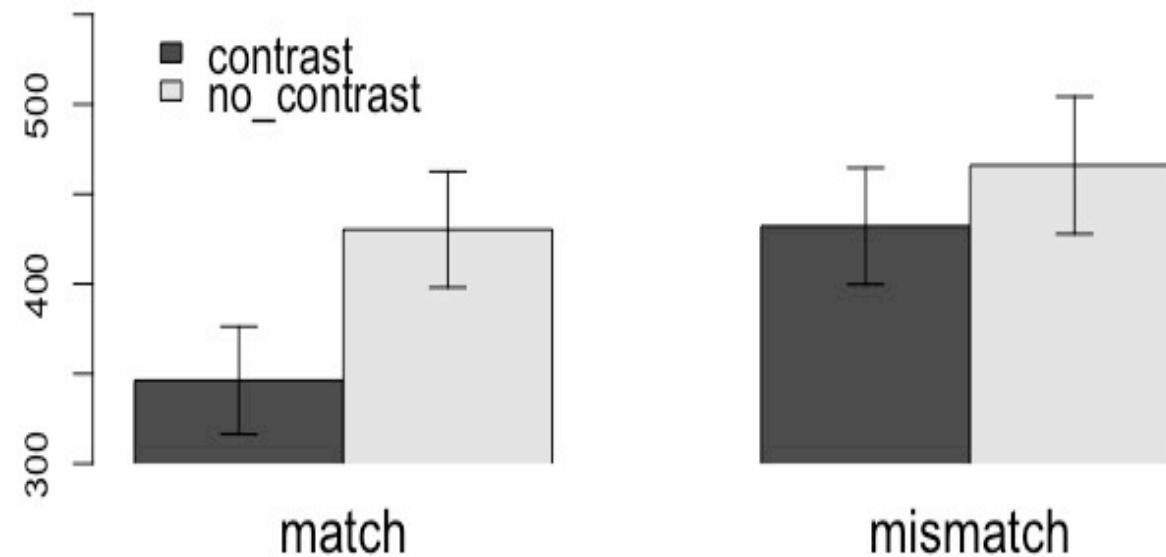
# First Pass Reading Times

| Fixed effects: | Estimate | Std. Error | df | t value | Pr(>|t|) |
|---|---|---|---|---|---|
| (Intercept) | 332.676 | 87.141 | 73.340 | 3.818 | 0.000279 *** |
| Matchmismatch | 85.423 | 40.839 | 50.290 | 2.092 | 0.041534 * |
| Contrastno_contrast | 83.850 | 35.856 | 40.380 | 2.338 | 0.024396 * |
| length | 1.480 | 8.285 | 69.320 | 0.179 | 0.858704 |
| Matchmismatch:Contrastno_contrast | -53.664 | 45.116 | 40.750 | -1.189 | 0.241141 |

# Log First Pass Reading Times

| Fixed effects: | Estimate | Std. Error | df | t value | Pr(>|t|) | |
|---|---|---|---|---|---|---|
| (Intercept) | 8.149257 | 0.299735 | 50.410000 | 27.188 | <2e-16 | *** |
| Matchmismatch | 0.306344 | 0.142958 | 39.620000 | 2.143 | 0.0383 | * |
| Contrastno_contrast | 0.306857 | 0.122778 | 42.970000 | 2.499 | 0.0163 | * |
| length | 0.009283 | 0.028246 | 44.180000 | 0.329 | 0.7440 | |
| Matchmismatch: | | | | | | |
| Contrastno_contrast | -0.223156 | 0.148470 | 34.200000 | -1.503 | 0.1420 | --- |
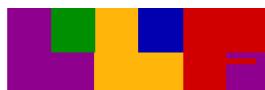
# Graph from course



**Regression path duration („go past")**

Claire knew the answer to the physics problem was wrong and decided ..
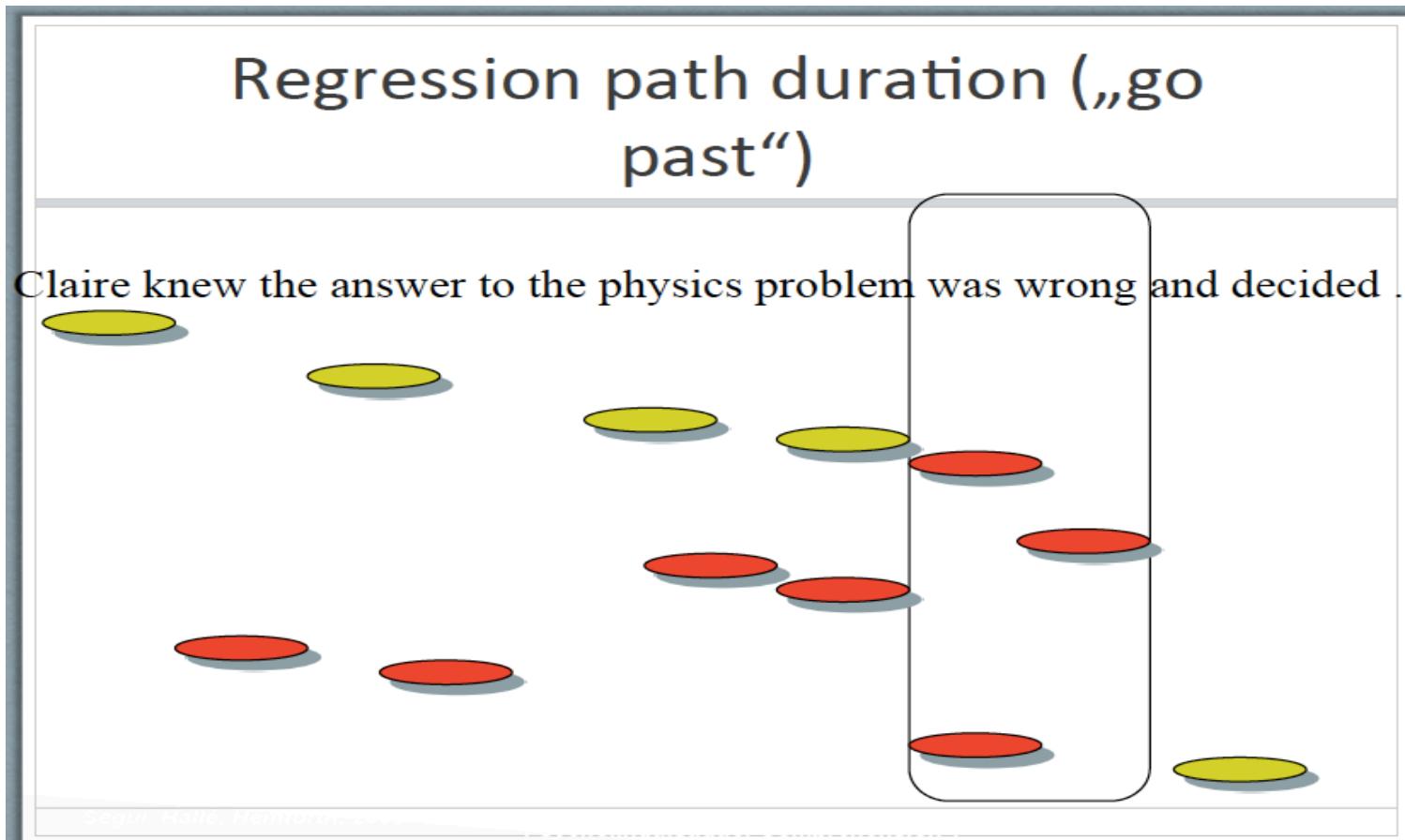
# Regression Pass Durations
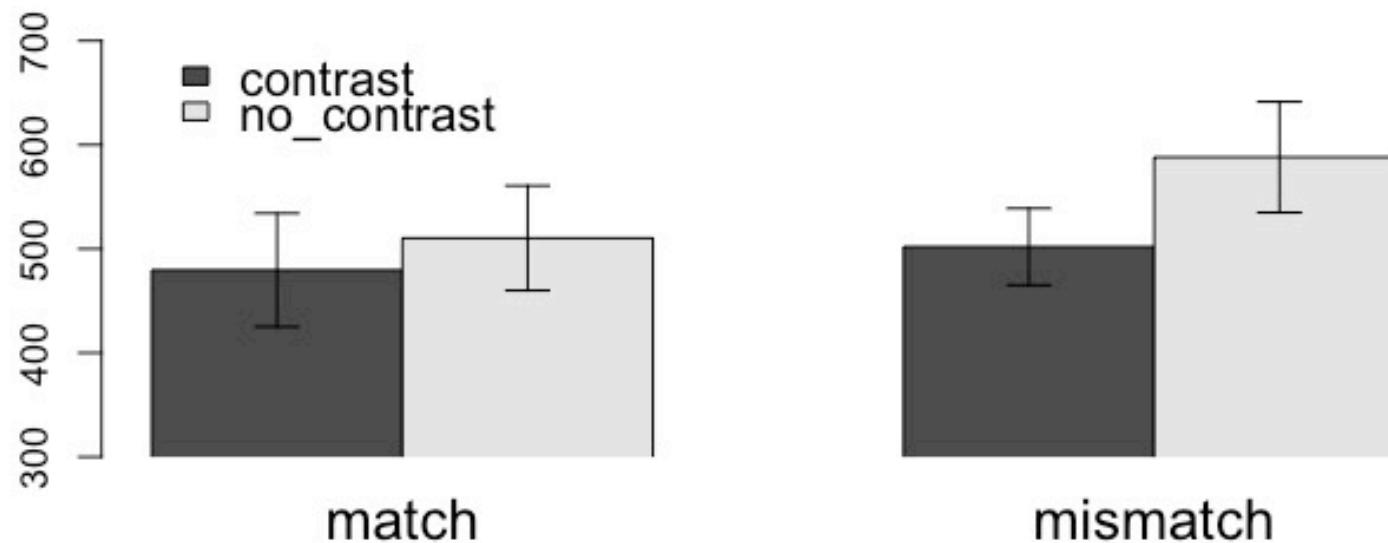
# Regression Path Durations "Minimal Model"

| Fixed effects: | Estimate | Std. Error | df | t value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| (Intercept) | 204.80 | 144.27 | 97.59 | 1.420 | 0.1589 |
| Matchmismatch | -65.96 | 56.87 | 160.18 | -1.160 | 0.2478 |
| Contrastno_contrast | 48.98 | 54.79 | 174.10 | -0.894 | 0.3726 |
| length | 27.30 | 13.96 | 90.19 | 1.956 | 0.0536 |
| Matchmismatch: | | | | | |
| Contrastno_contrast | 136.30 | 64.23 | 250.96 | 2.122 | 0.0348 * |

# Log Regression Path Durations Maximal Model

| Fixed effects: | Estimate | Std. Error | df | t value | Pr(>|t|) |
|---|---|---|---|---|---|
| (Intercept) | 8.08222 | 0.35960 | 97.03000 | 22.475 | <2e-16 *** |
| Matchmismatch | 0.05293 | 0.15884 | 73.58000 | 0.333 | 0.740 |
| Contrastno_contrast | 0.04874 | 0.16165 | 52.57000 | 0.301 | 0.764 |
| length | 0.04933 | 0.03388 | 86.41000 | 1.456 | 0.149 |
| Matchmismatch: Contrastno_contrast | 0.11835 | 0.18405 | 51.46000 | 0.643 | 0.523 |

# Graph from course



Total reading time
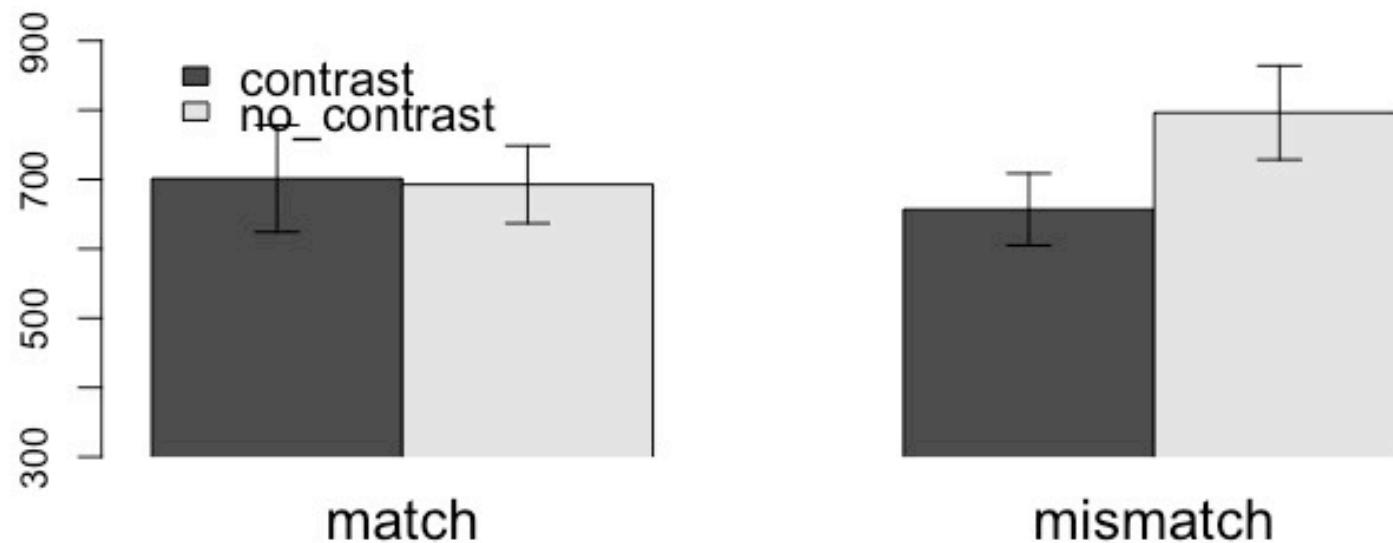
Claire knew the answer to the physics problem was wrong and decided ...

# Total Reading Times

# Total Reading Times "Maximal Model"

| Fixed effects: | Estimate | Std. Error | df | t value | Pr(>|t|) | |
|---|---|---|---|---|---|---|
| (Intercept) | 727.5741 | 204.2019 | 36.1400 | 3.563 | 0.00105 | ** |
| Matchmismatch | -51.9529 | 87.7834 | 46.2100 | -0.592 | 0.55685 | |
| Contrastno_contrast | -12.3277 | 88.1953 | 37.2500 | -0.140 | 0.88959 | |
| length | -0.9398 | 19.2698 | 30.8200 | -0.049 | 0.96142 | |
| Matchmismatch:Contrastno_contrast | 156.1445 | 101.6037 | 37.5400 | 1.537 | 0.13273 | |

# Log Total Reading Times "Maximal Model"

| Fixed effects: | Estimate | Std. Error | df | t value | Pr(>|t|) |
|---|---|---|---|---|---|
| (Intercept) | 8.944241 | 0.366387 | 32.580000 | 24.412 | <2e-16 *** |
| Matchmismatch | -0.065699 | 0.153962 | 37.430000 | -0.427 | 0.672 |
| Contrastno_contrast | 0.004302 | 0.157875 | 34.690000 | 0.027 | 0.978 |
| length | 0.022729 | 0.034404 | 27.220000 | 0.661 | 0.514 |
| Matchmismatch:Contrastno_contrast | 0.251766 | 0.181449 | 33.450000 | 1.388 | 0.174 |

---Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
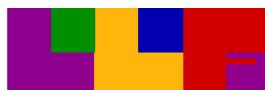
# Second Pass Times
## (only cases where the critical regions is read a second time)
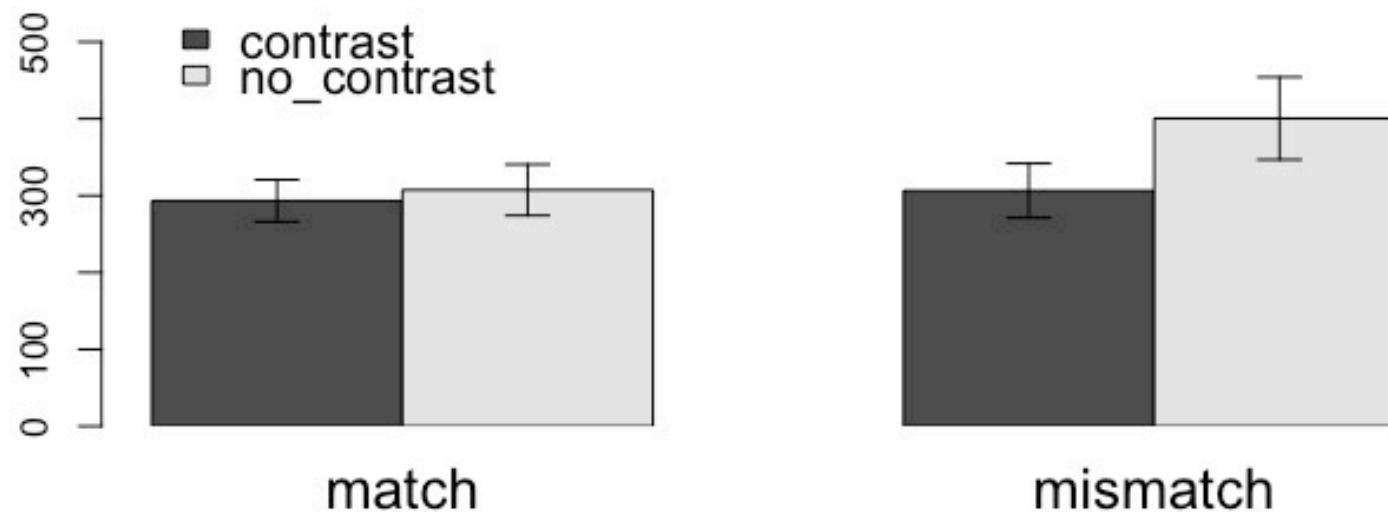
# Second PassTimes "Intercept Model" (not enough data for max model)

Fixed effects:        Estimate Std. Error        df        t value     Pr(>|t|)

(Intercept)          295.2181   104.3676 359.0000   2.829  0.00494 **

Matchmismatch      13.5585   44.9497 348.1000   0.302  0.76311

contrastno_contrast  13.5133    45.5413 347.6000   0.297  0.76685

length                    -0.1905    10.3785 350.8000  -0.018  0.98537

matchmismatch:contrastno_contrast  82.6597    51.6148 350.9000

1.601  0.11017
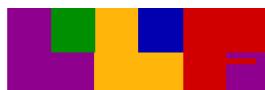
# Second PassTimes "Intercept Model", contrast within mismatch

| Fixed effects: | Estimate | Std. Error | df | t value | Pr(>\|t\|) | |
|---|---|---|---|---|---|---|
| (Intercept) | 360.808 | 209.065 | 22.030 | 1.726 | 0.09838 | . |
| contrastno_contrast | 96.791 | 33.324 | 163.550 | 2.905 | 0.00419 | ** |
| length | -4.066 | 15.551 | 22.430 | -0.261 | 0.79612 | |

# Log Second PassTimes "Intercept Model", contrast within mismatch

| Fixed effects: | Estimate | Std. Error | df | t value | Pr(>|t|) | |
|---|---|---|---|---|---|---|
| (Intercept) | 7.88403 | 0.84674 | 22.82000 | 9.311 | 3.1e-09 | *** |
| contrastno_contrast | 0.34003 | 0.12339 | 163.15000 | 2.756 | 0.00652 | ** |
| length | 0.01302 | 0.06292 | 23.23000 | 0.207 | 0.83792 | |

# Results

- – For First pass reading times, Mismatch and Contrast play a role.

- – For Second pass reading times, Mismatch is only hard without Contrast.

- – This may be true for Regression path durations and Total Reading Times (not yet tested)
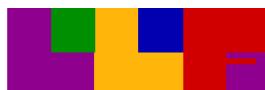
# 6 RNR analyses

Different approaches :

- ATB extraction (Ross 1967, Steedman 1996)

*[John likes - but Mary dislikes –] bananas.*

- multidominance (McCawley 1982, Moltmann 1992, Bachrach & Katzir 2008)

*[John likes but Mary dislikes] bananas.*

- phonological deletion (Kayne 1994, Hartmann 2000, Chaves 2014)

*[John likes ~~bananas~~] [but Mary dislikes bananas].*

# Formal analysis of Peripheral Ellipsis

Syntactic mismatch cast doubt on:

- syntactic raising or extraction

- multidominance

- deletion under identity

$\Rightarrow$call for a revised deletion identity

Abeillé, Crysmann, Shiraishi CSSP 2015

# Conclusions

- RNR is often considered to impose stricter identity conditions (syncretic forms)

- RNR with det, prep or voice mismatch attested  (in well written prose or official texts)

- Experiments show that RNR with det mismatch as acceptable as without; and voice mismatch (and semantic contrast) as acceptable as with voice match;

- Casts doubt on raising or multidominance analyses

=> For a revised deletion analysis in HPSG

# Prep mismatch in other languages

English (Bilbîie 2013)

*They were also as liberal or more liberal than any other age group in the 1986 through 1989 surveys* (WSJ,PTB)

*They were also as liberal as/*than any other age group …*

Spanish(Camacho 2003)

*Primero amedrentaron y luego dispararon contra los manifestantes.*

(First they-harassed and then they-shot at the demonstrators)

*Amedrentaron *(a) los manifestos*

Laboratoire de Linguistique Formelle

# Voice mismatch in other languages

## German

Einige haben sich von alleine, die anderen wurden im Inkubator künstlich vermehrt.

'Some have themselves alone, the others were in an incubator artificially multiplied.'

Einige haben sich gleich freiwillig, die restlichen wurden dann zwangsweise geopfert.

'some have themselves willingly, the rest have been by force sacrified'

# Voice mismatch in VP ellipsis

Kehler 2000: only with asymetric discourse relation

#This problem was looked into by John, and Bob did too.

Kertz 2014: more general condition that contrastive topics should be sentence topics=> voice mismatch with contrastive focus only

This information could have been released by Gorbachev, but he chose [not]F to. (Hardt 1993:37)

#The incident was reported by the driver, and [the pedestrian]T did too.

Peripheral ellipsis:

Symetric discourse relations ok with mismatch

Semantic contrast condition independent of mismatch

# English RNR and contrast

Subject contrast not enough

*Bill likes, and Mary likes, the TV show. (Ha 2008)

Corpus study by Bilbiie 2013 (Penn treebank)

VP (159) more common than S (41) coordination

Usually polarity, modality, aspect or tense contrast,

Did you or did you not say what I said you said... ? (brwn-4498)

But the South is, and has been for the past century, engaged in a wide-sweeping urbanization... (brwn-16897)

Who is and who should be making the criminal law here ? (wsj-6370)

Very few cases with different subjects:

...she knew and we knew that it was cowardice that had made one more radish that night just too impossible a strain. (brwn-5115)

The police said, all the people said, that's fine. (swbd-104656)