Summary meeting with Benjamin + Proposal
10-09-2015

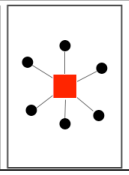
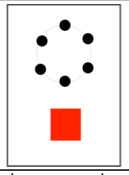
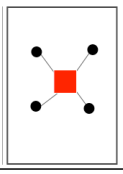
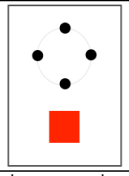
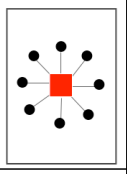
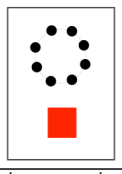
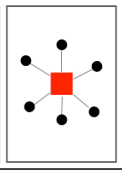
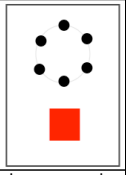
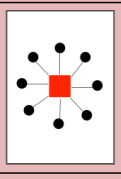
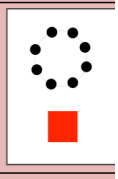
MAXIMALITY and PHANTOM READINGS – Priming experiment

Goal: Prime phantom readings by presenting sentences where non-maximal readings (non-upper-bounded) are clearly available (and maybe are the only possible interpretation).

What did we discuss?

- Decision about collective and distributive predicates (3 for each condition): Similar length, structure (generally, two arguments), and possibility of being depicted by same images (to reduce the difference between the two predicate types). See table below.
  [Note: We also discussed the possibility of presenting two different tasks for primes and targets; for instance, a reading task in primes, and a truth-judgment task in targets. This would allow having more variety of predicates in primes, which are not easy to draw. However, we ended by finding some predicates that could work very well.]
- Inclusion of fillers where maximal and non-maximal readings don't give rise to different truth-conditions (i.e. "more than three"). These can be used with both collective and distributive predicates.
- Although we didn't completely fix the task, we ended by thinking that a truth-judgment task might be simpler and informative in a first stage. If the binary-response is not informative enough (we get too many FALSE responses), we can do a graded judgment in a second step.

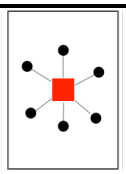
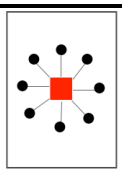
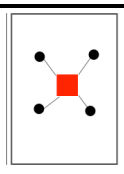
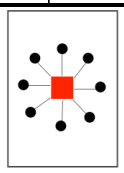
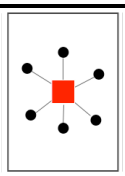
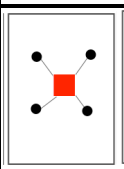
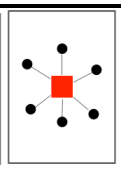
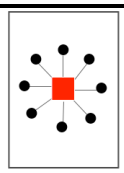
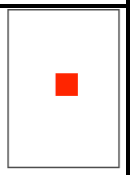
Examples of predicates and pictures for each type of trial

	PRIME				**FILLER**				**TARGET**	
	Between 5 and 7 dots…				**More than 7 dots..**				**Less than 7 dots…/ 5 dots**	
	Maximal reading (TRUE)		FALSE		TRUE		FALSE		??	
	6 dots		4 dots		8 dots		6 dots		8 dots	
										
Collective	surround a <SHAPE>	be arrange in <SHAPE>, form a <SHAPE>	surround a <SHAPE>	be arrange in <SHAPE>, form a <SHAPE>	surround a <SHAPE>	be arrange in <SHAPE>, form a <SHAPE>	surround a <SHAPE>	be arrange in <SHAPE>, form a	X	X
Distributive	be connected to the <SHAPE>	be above the <SHAPE>, be below the <SHAPE>	be connected to the <SHAPE>	be above the <SHAPE>, be below the <SHAPE>	be connected to the <SHAPE>	be above the <SHAPE>, be below the <SHAPE>	be connected to the <SHAPE>	be above the <SHAPE>, be below the <SHAPE>	be connected to the <SHAPE>	be above the <SHAPE>, be below the <SHAPE>
Baseline	There is a red square./There are black dots.		There are red dots.						X	X

[Note : Here, we use « Between m and n » in primes to have both false and true items]

PRE-TEST with graded judgments

We discussed the possibility of running a pilot / pre-test to be sure that, using those images, people are actually accessing to non-maximal readings in the expected degree. This would be an extension of what has been done by you with phantom readings for distributive predicates but including collective predicates.

PRE-TEST - GRADED JUDGMENTS									
	Between 5 and 7 dots…			More than 7 dots..			Less than 5/7 dots…		
Collective	Maximal reading	Non-Maximal reading	False?	Maximal reading	Non-Maximal reading	False?	Maximal reading	Non-Maximal reading	False?
Distributive	Maximal reading	Non-Maximal reading	False?	Maximal reading	Non-Maximal reading	False?	Maximal reading	Non-Maximal reading	False?
							 		

Note that this could be simplified, testing only non-maximal readings. In that case, we could do directly a baseline group (with targets + baseline trials) and see what does it happen in targets.

PRIMING EXPERIMENT WITH GLOBAL PRIMING

We think one possible design for a global priming experiment (with either groups or blocs of subjects) would be something like this:

TRIAL	MODIFIED NUMERAL	TRUTH-VALUE	GROUP			PREDICATE
			COLLECTIVE	DISTRIBUTIVE	BASELINE	
Filler	More than N	TRUE	[+] Distributive	[+]Collective	50-50	
		FALSE	[+] Distributive	[+]Collective	50-50	
Prime	Between M and N	TRUE	[+]Collective	[+]Distributive	[+]Baseline	
		FALSE	[+]Collective	[+]Distributive	[+]Baseline	
Target	Less than N	x2	[+] Distributive	[+]Distributive	[+]Distributive	

Basically, in all groups we have:

a. Same amount of distributive and collective predicates
b. Same amount of true-false items per predicate (i.e., target responses cannot be explained only in contrast with what participants do in primes)

Only one potential issue with the images, since there is one image that is always mapped with a False truth-value for critical sentences.

Decision:

- **Numerals (alternating or not)**
- **Shapes**
- **Mouse tracking**

Other points of discussion:

MOUSE-TRACKING AND PLURALS:

- We agreed on doing the experiment only with two predicates (connected and something else that changes the ratio).
- Potential influence of the task and the use of images: Maybe we should try something similar to what Tomlison et al did with implicatures, using either general knowledge or a short-history that gives such context.

BESIDES: Different possibilities with mouse-tracking (e.g. presupposition accomodation). Maybe simpler, that we know that we will have some kind of interesting result.

SUMMARY – Meeting 08-09-2015 (with Emmanuel)

1. MAXIMALITY and PHANTOM READINGS – Priming experiment

Goal: Prime phantom readings by presenting sentences where non-maximal readings (non-upper-bounded) are clearly available (and maybe are the only possible interpretation).

PRIMING TYPE X TASK
(To do: Explore which are the cognitive differences between global and local priming besides the difficulty to get each of them)

Task	Type of priming	
	GLOBAL	**LOCAL**
	Either by using group of subjects with only one type of prime o by using blocs with different primes.	**Trial-to -trial, just as the one used in our priming experiment last year**
Sentence-picture matching task (2 pictures, decision)	* Dependent variable: % non-maximal choices in targets * Prediction: Difference according to the previous prime or prime group (depending on type of priming) * What are we measuring? The alternative readings are given (by the images). If participants have a different interpretation, they might go with the "less worse" choice. * Potential problems: (1) using certain modified quantifiers (e.g., Less than) could make the sentence tautologically true (because if there is no maximality and no existential closure, for all possible quantities of dots the image will make the sentence true). (2) targets are very difficult to construct because if we don't propose a force choice (between a picture that makes the non-maximal reading true and a foil), we should make people choose between maximal and non-maximal readings.	
Sentence-picture matching task with interrogation/unknown possibility	* Dependent variable: % non-maximal choices in targets (vs. Some unknown possibility) * Prediction: Difference according to the previous prime or prime group (depending on type of priming). If the priming is with images that also make the upper-bounded reading true (see example), we expect priming only after collective primes; if the priming is with images that make all readings false, we expect priming for both primes (probably with a difference). * What are we measuring? The salient reading that participants actually have. If they have an alternative reading, they can go for the "uncertain" option. In those cases, we don't know which is the underlying interpretation. * Potential problems: We solve the problem in targets, but we have the problem of not knowing which reading people have.	
Truth-judgment task (TRUE-FALSE options)	* Dependent variable: % of TRUE responses * Prediction: Higher rate of true responses after collective primes or "false" primes (as above). * What are we measuring? The salient reading that participants actually have. If they have an alternative reading, they can go for the FALSE option. In those cases, we don't know which is the underlying interpretation. * Potential problems: Same as above.	
Graded truth-judgment task (continuous scale)	* Dependent variable: Continuous (percentages between 0 and 1) * Prediction: Increase of non-maximal rating after primes. * What are we measuring? The difficulty that participants have to access to the reading that the	

	image make true.

EXAMPLES of each trial x TASK

Trials	Possible tasks 1. Sentence-Picture matching task FORCED CHOICE	2. Sentence-Picture matching task WITH INTERROGATION	3. Truth-judment task YES-NO DECISION	4. Graded judgment task PERCENTAGE / 1-10
COLLECTIVE PRIME (NON-MAXIMALITY)	***Between 3 and 5 triangles form a square.***	Choosing the salient sentence reading. Choosing the alternative (and making salient whichever alternative)	TRUE FALSE	YES |_____| NO
DISTRIBUTIVE PRIME (MAXIMALITY)	***Between 3 and 5 dots are green.***	Choosing the salient sentence reading. Choosing the alternative (and making salient whichever alternative)	TRUE FALSE	YES |_____| NO
TARGET	***Between 1 and 3 hearts are red.*** The problem here is that target trials should be a free choice (preference), but choosing between maximal or non-maximal readings will potentially always give maximal readings.		TRUE FALSE	YES |_____| NO
BASELINE PRIME	***There are three triangles***			

** I included this option only to see the range of possibilities, although I think it's impossible to get the results that we expect with this configuration. First, we could be priming the 4-dots configuration from the beginning. Second, I guess the maximal reading is too strong to be chosen. However, if at some point we are interested on measuring priming effects with other measures, such as response times, maybe we do find an effect.

Discussion about one doubt of mine regarding maximality with indefinites (taken from Brisson; we decided not to pay attention to this now)[1].

2. MOUSE-TRACKING and PLURALS

A. Discussion about the properties of images and the low amount of distributive readings

Could the preference for cumulative readings (and the dispreferrence for distributive readings) be driven by the sentence-image relation? In order words, certain images could make much easier to derive one reading than the other (but importantly, other images might have the opposite pattern). Since in our experiment we are using the same images for both conditions, this type of bias could lead to (1) one reading at ceiling (as we do have), (2) low accessing rate for the other reading, and (3) an extra difficulty observed in RT and mouse trajectories for the dispreferred interpretation.

Which are the elements to support this idea?

1. **Scope results** suggesting that the images combined with particular quantifiers are driving most of the effect in response times and accessing (Note: it's not the case that only the images are driving the effect, otherwise in French we should have had the same results).
2. **Amount of distributive readings in different experiments:** Although it's true that the task is different in the priming and MT experiments, in the former the rate of distributive interpretations is higher than 50% (even when there is no priming). Besides the task itself, we didn't change many things: the images and the insertion of the adverb "exactly". Notice that the biggest change in images is a consequence of the inclusion of other predicates besides "be connected to".
3. **Predicate influence but no influence of numerals:** In fact, the results suggest that the rate of distributive responses rises when "be connected to" predicates to ~50%. Conversely, there is seems to be no effect of the number of shapes in the image. The maxRatio results per predicate also suggest that the difference between cumulative and distributive conditions is reduced in this predicate [See below].
4. **First priming experiment:** Our pattern of results in this MT experiment is very similar to the one that we had in the first pilot done for the priming experiment (distributive answers ~25%). In this case, the problem was the quantity of shapes and the disposition (i.e., a change in that gave us a 50-50 preference).

However:

a) **Differences with pilots in French.** We piloted a similar version of the experiments (same images) in French and results were as expected (50-50). Why in French is different? (See: the experiment was longer and only 10 participants).
b) Task might differentiate the results (it's not the same to be force to access the reading than to choose between that image and some other unknown possibility).

---

[1] Modified numerals allow non-maximal readings as phantom interpretations or as salient readings, depending on the type of predicate. This pattern seems to be inversed with definite expressions.

1. Distributive predicates
   a. The girls came. > Non maximal reading clearly available
   b. Between 2 and 5 girls came. > Non maximal reading = Phantom

Following Brisson, with certain collective predicates:
2. Collective predicates (depending the aspectual status)
   a. The girls are a big group. > Non maximal reading = Phantom or unavailable?
   b. Between 2 and 5 girls are a big group. > Non maximal reading clearly available?

Possible alternatives:

- To discard predicate effects: Change the predicates or the images for above/below
  Maintain one predicate 50-50 and one predicate as "above" to compare what happen when the rate of distributives decrease for image/predicate reasons.
- To discard task effects: Instead of using TRUE-FALSE responses, we could provide images. For example, we could have three types of trials: (1) common preference task with one image corresponding to a cumulative trial and another corresponding to a distributive trial; (2) one distributive image and one interrogation; (3) one cumulative image and one interrogation.
- Using context/history instead of images, to get a task more similar to the one used in Tomlison et al. (the task is easier with general knowledge).

Aside question: Are there experiments where the task is to create the pictures (people arrange shapes according to a sentence)? Revise

B. How to explain the high amount of deviation in both experimental conditions?

A baseline deviation towards the alternative might be associated to the existence of ambiguity itself, independently of the final reading (and the cost associated to each of them). Higher deviation is associated to distributive readings because the weight of cumulative readings is stronger, i.e. people access to a cumulative reading before accessing to a distributive reading.

In relation with this:

TOMLISON, BAILEY AND BOTT – Possible all of that and then some.

While in our results the high derivation associated to both conditions contributes to make mean trajectories very similar, their mean trajectories for "logical" and "pragmatic" (SI) conditions reflect very clearly two different patterns of derivation. How do we explain this difference?

In their experiments, T&B&B use one type of critical sentence (*Some elephants are mammals*), where logical interpretations = TRUE and pragmatic interpretations = FALSE. They have basically 4 control items (All-True, All-False, Some-True and Some-False). Notice that in Some-False trials (*Some elephants are insects*) both alternative readings are FALSE and in Some-True (*Some mammals are elephants*) both readings are TRUE. Neither of these conditions presents a big deviation towards the alternative.
When they compare for critical items True (41%) and False (59%) responses, they found that there is a difference of deviation. [Note: This effect is still there when they restrict the analysis to "pragmatic participants", who said false most of the time]

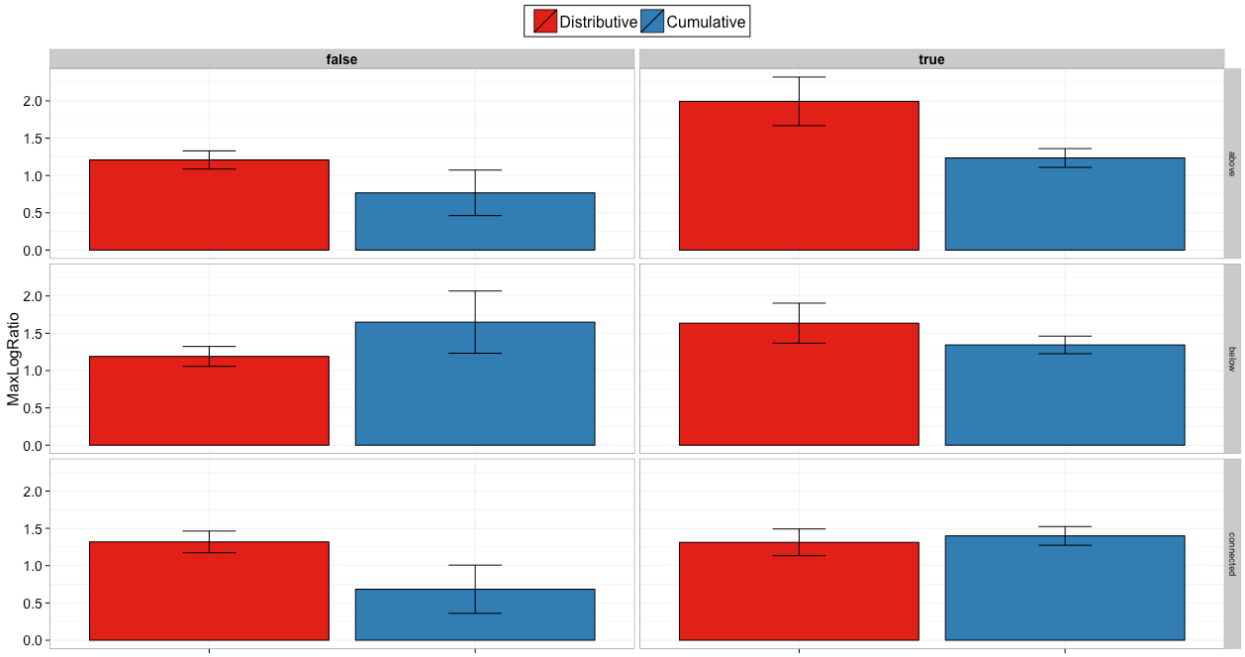
So, why our results are so noisy?

Until here with Emmanuel

This is independent of the type of derivation, whether it's in two-steps (somehow serial) or in parallel and cumulative readings are derived first because they are faster (cf. idea of Speed-Accuracy trade off).

Follow-up?

a. Modeling: Can the degree of noise explain by itself the difference between conditions?
b. Speed-Accuracy trade-off: Dissociate deriving faster one reading than other from two-step derivation. (I've explored this a little bit when Emmanuel proposed it)



UNTIL HERE WE DISCUSSED (OF MY HANDOUT)

---

3. MOUSE-TRACKING and SCOPE

Recap: Our results in English went in the opposite direction from what it was found in the literature (however, the task wasn't exactly the same one).

- Can we explain this effect as a consequence only of the images? If yes, how we understand our pilots.
  To-explore (Ewan- Alexis Wellwood, influence of images in semantic tasks)

- Exploit results in French. Using "tous" (all), we got the trend in pilots. Is it worthily to do a whole experiment in French? Maybe it's interesting to explore the cross-linguistic differences + the potential differences between quantifiers and their scope properties (cf. wide-scope preferences for Every in certain dialects of English)

4. PRIMING SCOPAL REPRESENTATIONS between QUANTIFIERS

March 2015: Pilot experiment (N=10) following the priming experiment by Chemla & Bott but testing the priming effect between quantifiers (i.e., 2-most, a-every and every-a). We also included different image types. In the results we observed a tendency of priming between representations for certain directions and we decided to do an experiment including only one type of image[2].

I programmed that experiment in that moment[3], including also fillers, but then we didn't run it because we were with too many things (I think).

Note: Complement of MT?

---

[2] Interestingly, when I was looking again at the results of this pilot experiment, I can see in primes a "trend" as the one that we found in our MT experiment; i.e., Inverse > Surface

[3] 192 trials, 144 experimental trials (72 primes+72 targets) and 48 fillers with sentences such as "There is a star in the middle" and "There is a square at the bottom" that reverse the responses expected for targets.

Other ideas for the future (not very thought, only if there is time):

5. Acquisition: Cumulativity, Distributivity and Collectivity

Syrett and Musolino (2013): Distributive vs. Collective (In terms of preference, they don't behave like adults). What about cumulativity? Why it could go in the opposite direction? How children deal with non-maximality issues?

6. Preference by cumulative/distributive readings under negation (English vs Spanish)

Context: There is a party and everyone is supposed to bring something to drink. Paul, Ann, Mary and John come.

Ex. (SP) Nadie trajo dos cervezas // (EN) No one brought two beers.

Different situations where the sentences would be true:

a. (Distributive TRUE and Cumulative TRUE) Ann brought one beer, and the others brought nothing.
b. (Distributive TRUE and Cumulative FALSE) Ann brought one beer, Mary brought one beer and John and Paul brought nothing.

Even though dissociating the readings is difficult, in Spanish the straightforward interpretation is (b) while in English apparently is (a) and (b) is difficult to get.

7. Exploring contextual and pragmatic influences in both priming and preference effects between distributive, cumulative and collective readings. For example, using big numbers can strengthen a cumulative interpretation (counting strategy for Distributive).