	Theme		Rheme		comments
	textual Theme	topical Theme	expected Theme		
1.		attack algorithms			
2.		L-BFGS			
3.		Szegedy *et al.* [46]		generated adversarial examples using box-constrained L-BFGS	
4.		given an image x	their method	finds a different image x' that is similar to x under L_2 distance yet is labeled differently by the classifier	Readers expect the Theme to be a Participant. In this clause, however, the Participant *their method* has been displaced by the Circumstance *given an image* x. It is by no means a mistake to displace the Theme expected by readers. Any time a reader encounters the unexpected, he or she will perk up and see what's going on. Nonetheless, for just this same reason, if a writer counters expectations every other clause, then (a) the reader will get worn down, and (b) the reader will lose orientation in the Message. Carlini and Wagner do an excellent job is expectation management, varying Themes always just often enough, not too much.
5.		they		model the problem as a constrained minimization problem [mathematical problem]	
6.		this problem		can be very difficult to solve however	

7.	so	Szededy *et al.*		instead solve the following problem [mathematical problem] where loss$_{F,1}$ is a function mapping an image to a positive real number	You may be wondering why the word *loss*$_{F,1}$ is not the Theme of a new clause. Clearly, the word *is* provided the base for a new clause, so why doesn't *loss*$_{F,1}$ take its place as Theme of that clause? The answer is this: **Not all clauses are created equal**. The clause around the verb *is* occurs inside of the other clause around the verb *solve*. As a result, the clause around the verb *is* loses status and becomes less salient in the Message. Technically, we say the clause is *downranked*, and this term nicely captures the idea that one clause gets swallowed by another. The most usual case of downranking can be viewed in Position 4 above: their method finds a different image *x'* downranks that is similar to *x* under L_2 distance You'll have learned this as a relative clause. So, the takeaway is, wherever you see a noun phrase (here, *a different image* x') that is followed by either *that* or *which*, there you can be certain of downranking. And what is so bad about downranking? Nothing really. But a writer should note that by using downranking, he or she removes the Theme of the downranked clause from the list of Themes contributing to the Message. For example, the noun phrase *a different image* x' belongs exclusively in the Rheme and will, accordingly, be removed from the Message. The downranking in this clause, Position 7, beginning at this word *where* is quite normal for mathematical problems. You can see another example of this kind of downranking in Position 15. On the other hand, you see further examples of the downranking of noun phrases in Positions 9, 16, 22, 30, and 31.

8.		one common loss function to use		is cross-entropy	
9.		line search		is performed to find the constant $c > 0$ that yields an adversarial example of minimum distance	
10.	in other words	we		repeatedly solve this optimization problem for multiple values of c, adaptively updating c using bisection search or any other method for one-dimensional optimization	I bet you've already noticed the first column is labeled *textual Theme*. And you've probably asked yourself what kind of tricks I'm trying to pull here. Well, I am pulling no tricks. There is such a thing as textual Theme. For example, in this clause, the phrase *in other words* is a textual Theme. The function of a textual Theme is to link clauses, and so the phrase *in other words* links this clause to the foregoing clause, at Position 9. The special thing about textual Themes is that they do not displace the real Theme of the clause. Basically, you can picture Thematicity as a drinking vessel, and the vessel only fills when a Participant or Circumstance as Theme is joined by one of these textual phrases. The Participant or Circumstance alone do not fill the vessel — there's still room under the brim for a linking phrase like *in other words*. The reason is, a textual Theme has two functions, one is to be thematic, but the overriding function is to link text. Therefore, a textual Theme functions less at the intra-clausal level, and more at the inter-clausal level Check out all the other examples of textual Themes down that column. See whether this makes sense to you, this joint thematic-and-textual function I'm explaining here.
11.		fast gradient sign			
12.		the fast gradient sign [11] method		has two key differences from the L-BFGS method	

13.	first	it		is optimized for the L_∞ distance metric	
14.	and second	it		is designed primarily to be fast instead of producing very close adversarial examples	This clause and the two previous are wonderfully clear. The clarity derives from the Theme of Position 12, *fast gradient sign*, repeating twice as *it*. Moreover, the way for the ultrashort Theme *it* is prepared for by the two textual Themes *first* and *second*. The effect of all this is to focus the Rheme, which is where the real point of these clauses resides. Carlini and Wagner write other such pairs of clauses between Positions 68 and 69, Positions 73 and 74, and Positions 75 and 76.
15.		given an image x	the fast gradient sign method	sets [mathematical problem] where ϵ is chosen to be sufficiently small so as to be undetectable, and t is the target label	
16.	intuitively	for each pixel	the fast gradient sign method	uses the gradient of the loss function to determine in which direction the pixel's intensity should be changed (whether it should be increased or decreased) to minimize the loss function	The parentheses de-emphasize content which otherwise would have provided more Theme and consequently, more topics. There is only one topic intended here: The means by which the fast gradient sign method minimizes the loss function. You can see two further examples of such use of parentheses at Positions 47 and 48. By the way, if you're interested in the function of the word *intuitively* in this clause, email me at daniel.shea@kit.edu
17.	then	it		shifts all pixels simultaneously	
18.		it is important to note			Technically, this is a simplification. But I'm not showing you this stuff to make you into a linguist. I'm showing you this stuff so that you understand text. Therefore, only the thing you need to note is this: Where you have *it* + *is* + evaluative adjective + *that*, there you really have just one Theme, and the meaning of that Theme is the evaluation by the adjective. In this case, the meaning of the Theme is *important*.

19.	that	the fast gradient sign attack		was designed to be *fast*, rather than optimal	
20.		it		is not meant to produce the minimal adversarial perturbations	
21.		iterative gradient sign			
22.		Kurakin *et al.*		introduce a simple refinement of the fast gradient sign method [26] where instead of taking a single step of size ϵ in the direction of the gradient-sign, multiple smaller steps α are taken, and the result is clipped by the same ϵ	
23.	specifically	begin		by setting [mathematical problem]	
24.	and then	on each iteration		[mathematical problem]	
25.		iterative gradient sign		was found to produce superior results to fast gradient sign [26]	
26.		JSMA			
27.		Papernot *et al.*		introduced an attack optimized under L_0 distance [38] known as the Jacobian-based Saliency Map Attack (JSMA)	

28.		we		give a brief summary of their attack algorithm	
29.		for a complete description and motivation	we	encourage the reader to read their original paper [38]	
30.		at a high level	the attack	is a greedy algorithm that picks pixels to modify one at a time, increasing the target classification on each iteration	
31.		they		use the gradient [mathematical definition] to compute a *saliency map*, which models the impact each pixel has on the resulting classification	
32.		a large value		indicates	
33.	that	changing it		will significantly increase the likelihood of the model labeling the image as the target class *l*	
34.		given the saliency map	it	picks the most important pixel	
35.	and			modifies it to increase the likelihood of class *l*	
36.		this		is repeated	
37.	until either	more than a set threshold of pixels		are modified	

38.		which		makes the attack detectable	This may look like downranking, but it's not. How can you tell the difference? Well notice how here the word *which* does not just pick up on a preceding noun phrase. The word which does not refer just to *pixels* or *threshold*. No, the word *which* is actually picking up on the entirety of Position 37; that is, the attack becomes detectable **after the occurrence of that entire action of very many pixels undergoing modification**.
39.	or	it		succeeds in changing the classification	
40.		in more detail	we	begin by defining the saliency map in terms of a pair of pixels p, q	
41.		define		[mathematical definition]	
42.	so that	α_{pq}		represents	
43.		how much changing both pixels p and q		will change the target classification	
44.		β_{pq}		represents	
45.		how much changing both pixels p and q		will change all other outputs	
46.	then	the algorithm		picks [mathematical problem]	
47.	so that	α_{pq}		> 0 (the target is more likely)	
48.		β_{pq}		< 0 (the other classes become less likely)	

49.	and	$\alpha_{pq} \cdot \beta_{pq}$		is largest	
50.		notice			The Theme here draws all the reader's attention to one single point: **procedure in JSMA**. Notice how such a purely thematic clause as this has the same effect on readers as does the headings of the section and the subsections at Positions 1, 2, 11, 21, 26, and 66.
51.	that	JSMA		uses the output of the second-to-last layer *Z*, logits, in the calculation of the gradient	
52.		the output of the softmax *F*		is *not* used	
53.		we		refer to this as the JSMA-Z attack	
54.	however, when	the authors		apply this attack to their defensively distilled networks	
55.		they		modify the attack	
56.	so	it		uses *F* instead of *Z*	
57.	in other words	their computation		uses the output of the softmax (*F*) instead of the logits (*Z*)	
58.		we		refer to this modification as the JSMA-F attack	
59.	when	an image		has multiple color channels (e.g., RGB)	

60.		this attack		considers the L_0 difference to be 1 for each color channel changed independently	
61.	so that if	all three color channels of one pixel change		change	This is a wonderful way to preload all the relevant conditions so that the really important point about the L_0 norm can be made alone in the Rheme of the next clause.
62.		the L_0 norm		would be 3	
63.	while	we		do not believe	
64.		this		is a meaningful threat model	
65.	when	comparing to this attack	we	evaluate under both models	
66.		Deepfool			
67.		Deepfool [34]		is an untargeted attack technique optimized for the L_2 distance metric	
68.		it		is efficient	
69.	and			produces closer adversarial examples than the L-BFGS approach discussed earlier	
70.		the authors		construct Deepfool by imagining	
71.	that	the neural networks		are totally linear, with a hyperplane separating each class from another	

72.	from this	they		analytically derive the optimal solution to this simplified problem	
73.	and			construct the adversarial example	
74.	then, since	neural networks		are not actually linear	
75.		they		take a step towards that solution	
76.	and			repeat the process a second time	
77.		the search		terminates	
78.	when	a true adversarial example		is found	
79.		the exact formulation used		is rather sophisticated	
80.		interested readers		should refer to the original work [34]	

commentary

Here are the large tendencies of Carlini and Wagner's text above:

(1) The Theme is consistently shorter than the Rheme.

(2) The Rheme is consistently the real point that the clause is making.

(3) Accordingly, at that stage in the discourse anyway, the Theme is consistently more familiar than is the Rheme.

(4) The choice of Theme seldom defies the expectations of readers; in other words, the column *expected Theme* is usually unoccupied.

(5) The Theme is consistently a Participant, and that Participant will repeat over a few clauses. Consequently, a change in Participant announces a change in Theme announces a change in topic. All of this is very orderly and thus helps orient readers in the discourse.

All these large tendencies are built into every next clause and into every next choice of Theme. Therefore, the big achievement in Carlini and Wagner's prose is made in their small choices at the level of the clause.

Carlini and Wagner work with expectations by generally meeting expectations.