Project: spaCy

2016-06-19T16:11:15Z blang 24comments

pipe(): ValueError Error parsing doc

Web url: https://github.com/explosion/spaCy/issues/429

API url: https://api.github.com/repos/explosion/spaCy/issues/429

I found strange behaviour using the `pipe()` method (only verified on german variant):

If you parse a document using `pipe()` you can get a ValueError, while if i use `nlp(text)` everything is fine. I boiled it down to single words, while german words work, english words like 'windows' don't work.

Steps to reproduce:

``` python

import spacy

nlp = spacy.load('de')

def texts():

yield "Windows"

for doc in nlp.pipe(texts(), n\_threads=16, batch\_size=1000):

print(len(doc)) # doc access -> ValueError

```

Trace

```

ValueError Traceback (most recent call last)

<ipython-input-2-9a095ec5505b> in <module>()

8 def texts():

9 yield "Windows"

---> 10 for doc in nlp.pipe(texts(), n\_threads=16, batch\_size=1000):

11 print(len(doc))

.../venv/lib/python3.4/site-packages/spacy/language.py in pipe(self, texts, tag, parse, entity, n\_threads, batch\_size)

254 stream = self.entity.pipe(stream,

255 n\_threads=1, batch\_size=batch\_size)

--> 256 for doc in stream:

257 yield doc

258

ValueError: Error parsing doc: Windows

```

If you use `nlp("Windows")` it works fine. Also if you execute `nlp("Windows")` before the same `pipe()` call, `pipe()` does not raise an exception (a dictionary is built?)

Versions:

```

Python 3.4.3 (Problem not related to ipython)

spacy 0.101.0

```

Maybe this is related to this region [syntax/parser.pyx](https://github.com/spacy-io/spaCy/blob/1822bb4ff1e6ceee76fe5e43de4345fab168e65b/spacy/syntax/parser.pyx#L179-L183)

``` python

if not eg.is\_valid[guess]:

# with gil:

# move\_name = self.moves.move\_name(action.move, action.label)

# print 'invalid action:', move\_name

return 1

```

-------------------------------------------------------------------------

2016-06-22T10:19:57Z pdeburen

Same issue here with the german model and `nlp.pipe()` on Amazon Linux (also on Ubuntu Server 14.04 LTS) using python 3.5.

However, blang's minimal example works on my Macbook (OSX 10.11.3) where I don't have OpenMP support in place (obviously only in single-thread). Setting `n\_threads=1` on Linux doesn't solve the issue for me.

-------------------------------------------------------------------------

2016-08-11T10:00:26Z pdeburen

I just tried this again and it seems to work now (reinstalled spaCy and the german model). @blang can you confirm? @syllog1sm Out of curiosity: Has there been an update to the German model which fixed this? Or was it a code change?

-------------------------------------------------------------------------

2016-08-18T14:52:52Z ontoport

Same problem here, working on Windows 10 with German text. Thought it was German that made it break. I also reinstalled spaCy and the German model yesterday, but this din't fix the problem in my case. I then tried to break it down to a specific sentence, but even after having removed this and succesively the follwoing sentences from my texts, the problem remained the same.

As above, if I use nlp(text) everything is fine.

-------------------------------------------------------------------------

2016-09-03T09:00:39Z elyase

I also had this problem with english text, it looks like a parser issue. Steps to reproduce:

``` python

def texts():

yield "11th September 11 years ago I started my first business"

for doc in nlp.pipe(texts()):

pass

```

raises:

```

---------------------------------------------------------------------------

ValueError Traceback (most recent call last)

<ipython-input-66-a423c90922c1> in <module>()

3 yield "11th September 11 years ago I started my first business"

4

----> 5 for doc in nlp.pipe(texts()):

6 pass

/Users/elyase/miniconda3/envs/intent/lib/python3.5/site-packages/spacy/language.py in pipe(self, texts, tag, parse, entity, n\_threads, batch\_size)

254 stream = self.entity.pipe(stream,

255 n\_threads=1, batch\_size=batch\_size)

--> 256 for doc in stream:

257 yield doc

258

/Users/elyase/miniconda3/envs/intent/lib/python3.5/site-packages/spacy/syntax/parser.pyx in pipe (spacy/syntax/parser.cpp:7143)()

ValueError: Error parsing doc: 11th September 11 years ago I started my first business

```

Curious thing, If you add a comma like this:

```

"11th September, 11 years ago I started my first business"

```

the error goes away. Directly doing:

```

doc = nlp("11th September 11 years ago I started my first business")

```

is ok in both cases.

-------------------------------------------------------------------------

2016-09-27T17:08:51Z honnibal

I think the `.pipe()` method isn't actually to blame here ‚Äî rather, it reports on an error condition thrown by the `parseC` method, while the `\_\_call\_\_` method fails silently in the same situation. I've fixed the silent failure, so that errors will be reported more consistently, but the underlying issue is possibly tricky.

The issue is arising because the entity recogniser's push-down automaton finds itself in a state with no continuations. I haven't stepped through the automaton yet (if you want to do that, use the method`nlp.entity.step\_through(doc)`) to see exactly where the problem is, but I'm pretty sure the error will come from an interaction with entities pre-set by the `Matcher` class. In order to preserve these entities, we restrict the actions of the entity recogniser, so that it can't over-write the previous ones. There's apparently a bug in the logic to introduce this constraint, that's leaving the automaton with no available actions. This results in an invalid predicted action, leading the parser to return a status code (it can't raise an error, as it's in a `nogil` function). This is the status code the `\_\_call\_\_` method was ignoring.

I'm afraid that 1.0 might paper over this problem, because the matcher won't by default set entities anymore ‚Äî this will be up to the user's control (there's a better API for customising the pipeline, though).

-------------------------------------------------------------------------

2016-10-07T16:55:21Z mattilyra

This might have something to do with state between the documents (don't know what if any is kept). I was having this exact issue with the german model and decided to just randomly shuffle the corpus to see what happens. At first errors were still being produced, but after a few shuffles of the corpus to errors to my surprise went away.

I'll try to produce a more reliable report of the behaviour.

\*\*update\*\* so the error going away didn't have anything to do with `random.shuffle`. However if you take the document that produces the error, pass it through `nlp(doc)` and then call `.pipe([doc])` again, it works.

An example document where `9/11` causes the error, removing that one token everything is fine.

```

doc = """Nichtsdestoweniger sagen wir Ihnen zu, dass wir genauso wie bei OEF daran arbeiten werden, dass dieses Mandat eine andere Struktur erh√§lt, und zwar aus politischen Gr√ºnden, weil zwar nicht rechtlich, aber nat√ºrlich politisch zur Kenntnis zu nehmen ist, dass 9/11 schon einige Zeit her ist."""

```

Tested on `0.101` and `master` - the issue seems to be fixed on `master` not just for the example document above, but for the entire corpus where that doc came from.

-------------------------------------------------------------------------

2016-10-19T00:34:29Z honnibal

I think I have this taken care of, but I'm not 100% sure. Please reopen if it reoccurs.

-------------------------------------------------------------------------

2016-10-27T13:00:04Z sadovnychyi

FYI it did happen for me with 1.1.0, but so far I cannot provide any steps to reproduce it.

The text it tried to parse isn't relevant: `Waiver of Jury Trial` but I did update global `nlp.matcher` in a loop while parsing the doc in the same loop. I'll get back to this if I'll be able to reproduce it with specific steps.

PS: Now I'm getting `Segmentation fault: 11` and I have no idea if it's relevant or not..

-------------------------------------------------------------------------

2016-10-27T13:05:23Z honnibal

Do you have a minute to video chat about this? If so click here:

https://appear.in/spacy\_issue429

-------------------------------------------------------------------------

2016-10-27T13:07:34Z sadovnychyi

Sorry, my internet isn't good for video chatting, but I'm happy to text.

-------------------------------------------------------------------------

2016-10-27T13:11:26Z honnibal

No worries.

If you're getting a segfault the handiest thing to do would be to break out the pipeline manually. Instead of:

```

doc = nlp(text)

```

You can do:

```

doc = nlp.tokenizer(text)

nlp.tagger(doc)

nlp.parser(doc)

nlp.entity(doc)

matches = nlp.matcher(doc)

# Act on your matches

```

Then you can investigate what's going on.

-------------------------------------------------------------------------

2016-10-27T13:14:52Z sadovnychyi

The segfault is caused by matcher. The number of matches I have is up to a million, python process eats about 4 GB of ram, and there's still enough for it to grow. I could investigate this later, maybe in another issue.

Trying to narrow the scope of ParserStateError right now.

-------------------------------------------------------------------------

2016-10-27T13:16:45Z honnibal

Hmm. Is the match proliferation expected for your use-case?

-------------------------------------------------------------------------

2016-10-27T13:36:26Z sadovnychyi

``` python

import spacy

from spacy.attrs import ORTH

nlp = spacy.load('en')

def merge\_phrases(matcher, doc, i, matches):

if i != len(matches) - 1:

return None

spans = [(ent\_id, label, doc[start:end]) for ent\_id, label, start, end in matches]

for ent\_id, label, span in spans:

span.merge('NNP' if label else span.root.tag\_, span.text, nlp.vocab.strings[label])

doc = nlp('a')

nlp.matcher.add('key', label='TEST', attrs={}, specs=[[{ORTH: 'a'}]], on\_match=merge\_phrases)

doc = nlp('a b')

```

->

```

Traceback (most recent call last):

File "case.py", line 17, in <module>

doc = nlp('a b')

File "/usr/local/lib/python3.5/site-packages/spacy/language.py", line 313, in \_\_call\_\_

proc(doc)

File "spacy/syntax/parser.pyx", line 117, in spacy.syntax.parser.Parser.\_\_call\_\_ (spacy/syntax/parser.cpp:6104)

spacy.syntax.parser.ParserStateError: Error analysing doc -- no valid actions available. This should never happen, so please report the error on the issue tracker. Here's the thread to do so --- reopen it if it's closed:

https://github.com/spacy-io/spaCy/issues/429

Please include the text that the parser failed on, which is:

'a b'

```

re: Hmm. Is the match proliferation expected for your use-case?

It won't grow much after this, I'm just curious how much entities it can hold and how it will affect the memory and performance. Should I open another issue for that segfault?

```

>>> spacy.about.\_\_version\_\_

'1.1.2'

```

-------------------------------------------------------------------------

2016-10-27T13:45:02Z honnibal

> It won't grow much after this, I'm just curious how much entities it can hold and how it will affect the memory and performance. Should I open another issue for that segfault?

I can easily make the matches list a numpy array if necessary.

A segfault via the Python API (as opposed to the Cython API) is always a bug. So yes, please open an issue.

-------------------------------------------------------------------------

2016-10-27T15:36:12Z sadovnychyi

I'll do it tomorrow, once I know the steps to reproduce it.

I guess now you have enough info for bug related to current issue.

On Thu, Oct 27, 2016, 21:45 Matthew Honnibal notifications@github.com

wrote:

> It won't grow much after this, I'm just curious how much entities it can

> hold and how it will affect the memory and performance. Should I open

> another issue for that segfault?

>

> I can easily make the matches list a C++ vector if necessary.

>

> A segfault via the Python API (as opposed to the Cython API) is always a

> bug. So yes, please open an issue.

>

> ‚Äî

> You are receiving this because you commented.

> Reply to this email directly, view it on GitHub

> https://github.com/explosion/spaCy/issues/429#issuecomment-256644729,

> or mute the thread

> https://github.com/notifications/unsubscribe-auth/AAL1SKptmaW0BE-yVBdMd5POkfwfMSTQks5q4KrkgaJpZM4I5Mhy

> .

-------------------------------------------------------------------------

2016-10-27T15:53:59Z honnibal

Yes. I have the test set up and I'm pretty sure I understand the problem now. Fix should be out soon.

-------------------------------------------------------------------------

2016-10-27T16:27:10Z honnibal

I think this should fix the segfault too ‚Äî I think they were related.

Closing for now. Again, if it reoccurs, don't hesitate to reopen :)

-------------------------------------------------------------------------

2016-12-31T01:38:38Z paulkarpenko

UPDATE: I was able to get around this by converting multiple spaces to a single space. Not sure if this was an issue with my string or with spaCy's processing.

@honnibal A spaCy error told me to reopen this thread :-/ Not sure I have the rights to do that, but here's the text:

```

File "spacy/syntax/parser.pyx", line 146, in spacy.syntax.parser.Parser.\_\_call\_\_ (spacy/syntax/parser.cpp:6114)

spacy.syntax.parser.ParserStateError: Error analysing doc -- no valid actions available. This should never happen, so please report the error on the issue tracker. Here's the thread to do so --- reopen it if it's closed:

https://github.com/spacy-io/spaCy/issues/429

Please include the text that the parser failed on, which is:

'The 2016 election has been all about Donald Trump ‚Äî even down ballot. His campaign has had a huge effect on the battle for the House, driving the two-year string of events that led us to Election Day. Here are the 16 moments that defined the 2016 race for the House. 1. House Republicans stretch their majority to historic proportions in 2014If Republicans hadn‚Äôt expanded their majority to a whopping 30 seats in 2014, who knows what position they would be in today. But by winning their biggest majority since 1928 that year, the GOP stretched Democrats‚Äô resources over a large 2016 battlefield packed with incumbents, who largely ran strong campaigns. Even at their best moments this campaign, Democrats have been reluctant to say they could challenge for the majority, and the 2014 election is a big reason why. 2. Judges torpedo Florida‚Äôs congressional mapOne reason Democrats are sure to make gains in the House this year is that courts in several states ordered re-draws of several gerrymandered congressional maps. Florida‚Äôs was the biggest: The new version of the map turned one Democratic district Republican, but it also tilted two GOP-held districts decisively Democratic, made a Republican-held battleground seat more Democratic-leaning, and put another veteran GOP lawmaker in a swing district for the first time. Democrats also gained a seat from court-ordered redistricting in Virginia, while several incumbents in both states and North Carolina lost primary challenges because they were put in new territory. 3. Trump‚Äôs Super Tuesday win fuels Democratic recruiting surgeThough Democrats have mostly been realistic about their chances of winning the House majority this year, Trump‚Äôs success in the GOP presidential race brought with it a surge of optimism ‚Äî and new candidates. Starting in early April, a handful of Democrats who would go on to become real threats to beat GOP incumbents launched late House campaigns in Colorado, Florida, Kansas, and more. A number of them, like Minnesota‚Äôs Terri Bonoff, cited Trump as a factor in their decisions to run ‚Äî and later made him a focus of their advertising against local incumbents. 4. The House Freedom Caucus takes John Boehner‚Äôs congressional districtThe hard-line House Freedom Caucus helped force Speaker John Boehner‚Äôs resignation in 2015, and then it added a symbolic insult to the establishment‚Äôs injury in March, when HFC-endorsed candidate Warren Davidson won the GOP primary to take over Boehner‚Äôs seat. The Freedom Caucus is sure to have continued influence in the House GOP Conference ‚Äî with its numbers staying steady around 40 while the overall size of the Republican majority shrinks ‚Äî and Davidson‚Äôs win was a big wakeup call to establishment donors that they were getting caught sleeping in the primaries that determined the makeup of their majority. 5. Democrats leave districts on the table in the springWhile Trump did help the party turn more House seats into battlegrounds this fall, Democrats also failed to get viable candidates in some districts that could have gotten interesting. An April primary in the Philadelphia suburbs ended in embarrassing defeat for one DCCC-backed candidate, while the victor has raised little money since. Other seats in New Mexico, southeastern Virginia, New Jersey, and more ‚Äî where Democrats have made strong challenges in the past ‚Äî fell by the wayside without strong candidates. 6. Donald Trump becomes the presumptive Republican presidential nomineeIt all starts with Trump. More than any other members of their party except Hillary Clinton, House Democrats have tried to make their local campaigns about the GOP nominee. Trump has fueled an ‚Äúeducation gap‚Äù between the parties, which has endangered Republican House members in white-collar suburbs that were once safer for the GOP ‚Äî but also shored up Republicans in blue-collar swing districts in the Northeast and Midwest. Veteran incumbents on both sides (including GOP Reps. Darrell Issa outside San Diego and John Mica outside Orlando, as well as northeastern Minnesota Democrat Rick Nolan) could fall victim to this trend. 7. Paul Ryan delays on endorsing TrumpThe House speaker eventually gave the GOP presidential nominee his support, but Ryan‚Äôs delay set the stage for a half-year of separation from the top of the ticket. Ryan‚Äôs touting of a ‚ÄúBetter Way‚Äù helped give House Republicans cover and maneuverability with regard to Trump, as many sought to out-run him by significant margins in their districts to win reelection. Ryan also helped GOP candidates in a big way by raising more than $35 million for the NRCC as part of over $50 million overall to protect the Republican majority. 8. California (and Latinos) sprint from the GOPWhile Hillary Clinton no doubt would have liked to wrap up the Democratic presidential nomination long before California‚Äôs June 7 primary, the late contest between her and Bernie Sanders fueled a voter registration surge in the state that may pay huge dividends in the fall. After initially looking to gain seats in California this year, House Republicans are now on defense in at least three toss-up districts in the state, where about 2.3 million new voters registered before the primary, including large numbers of Democratic-leaning Latinos and young voters. That state alone could put a dent in the GOP majority on Nov. 8, and increasing Latino turnout in Florida, Nevada, and several other states could net yet more Democratic districts, too. 9. Bernie Sanders makes his presence felt down-ballotWhile Sanders‚Äô presidential campaign fell short, the donor base he established became a huge boon for some House Democrats. Starting in April, Sanders has endorsed a handful of candidates around the country and raised hundreds of thousands of dollars for them by emailing his fundraising list on their behalf. That‚Äôs huge money in House races, especially primaries. It helped some candidates win nominations and made others competitive where otherwise they would not have been able to run TV ads. And in the fall, Sanders-raised money helped plug holes in some House Democrats‚Äô budgets when GOP attacks arrived in their districts before backup from Democratic outside groups. 10. Roger Marshall defeats HFC member Tim Huelskamp in a Kansas primaryUsually it‚Äôs the tea party that unseats GOP incumbents in primaries, but challenges to veteran Reps. Bill Shuster, Kevin Brady, and more fell short in 2016. Meanwhile, Kansas physician Roger Marshall beat GOP Rep. Tim Huelskamp by arguing that Huelskamp‚Äôs tea-party positions had made him an ineffective advocate for the state‚Äôs agriculture-heavy ‚ÄúBig First‚Äù district. With backup from the U.S. Chamber of Commerce and several ‚Äúestablishment‚Äù-aligned super PACs, Marshall will take over the safe Republican seat in January. The outside groups that backed him also helped defeat potential Freedom Caucus members in two late, open GOP primaries. But it may also provoke a response from the Freedom Caucus and other conservative groups who now feel targeted by establishment donors. The 2018 GOP primaries may come with fireworks. 11. The DCCC uses Trump to send extra cash to candidatesOne side effect of Democrats‚Äô late-breaking House optimism this year was that some of their most important candidates were low on funds, whether because they started campaigns late or donors were slow to take notice. So around the end of summer and beginning of fall, those candidates started airing specially worded TV ads connecting local candidates to national Republicans and Donald Trump ‚Äî and splitting hundreds of thousands of dollars each in advertising costs with the DCCC. By the last week of the election, the DCCC and at least 29 candidates had split over $14.2 million in coordinated advertising costs. The Democratic committee has not explained its legal reasoning on the Trump-focused ads, but the end result is that the party sent millions of extra dollars to candidates who might otherwise have not been able to run viable campaigns. 12. Trump‚Äôs 2005 Access Hollywood tape publishesAfter tape leaked of Trump making vulgar comments about sexually assaulting women, it looked like the bottom might fall out for down-ballot Republicans. Democratic groups began pushing money into new districts, and House Republicans foresaw not only a collapse in Trump‚Äôs numbers (which continued through the debates) but the potential that their base would stay home on Election Day. It was a doomsday scenario that had Democrats and Republicans alike suddenly scanning the horizon for a wave. 13. Republican super PACs douse potential fires with late moneyWhen things started to look bad for the congressional GOP in early October, House Republicans‚Äô biggest super PAC was there to fight the fire. Congressional Leadership Fund and its sister nonprofit, American Action Network, dumped $10 million into races where Democrats could have expanded the map right after the Access Hollywood tape broke, giving cover and a financial advantage to some Republican incumbents worried the national tide was about to turn against them. 14. Obama goes all-inPresident Barack Obama has not played a big role for individual Democratic House candidates while in the White House. In 2012, he was concentrating on his own reelection, and he was not popular in swing districts during the 2010 and 2014 midterms. But Obama made endorsements and cut TV and radio ads in dozens of districts in mid-October, looking to push the House in Democrats‚Äô direction as part of his political legacy. Keen observers took note that the president‚Äôs first House ad of the general election backed Illinois Democrat Brad Schneider, a candidate who had opposed the White House‚Äôs nuclear deal with Iran during his primary (before grudgingly coming around). It was a welcome sign for House Democrats looking to make final pushes against better-known incumbents. 15. GOP starts ‚Äúcheck and balance‚Äù strategy against Clinton and House DemocratsAs Trump‚Äôs prospects faded in October, Republican candidates and committees increasingly began running ads basically admitting that Clinton would win the White House ‚Äî and arguing that voters should elect Republican House members to keep her in check. The ads weren‚Äôt as widespread as in 1996, when congressional Republicans largely abandoned Bob Dole at the end of the presidential race. But Clinton‚Äôs popularity was never high this election, and Republicans believed the strategy would help them hold districts against Democratic candidates about whom voters knew relatively little. 16. The FBI announces it is still looking into Clinton-related emailsAn October surprise and bookend to the Trump tape, the FBI‚Äôs late email announcement helped re-energize the Republican base and, strategists believed, foreclosed on that possibility of bad turnout because of Trump‚Äôs early-October struggles. It also halted House Democratic momentum just as the party was trying to expand the battleground map late.

-------------------------------------------------------------------------

2017-01-20T07:38:52Z muthupl

Hi,

We are getting a parser state error. Here is the trace:

Traceback (most recent call last):

File "tests/test\_spacy\_nlp.py", line 231, in test\_should\_return\_none\_when\_spacy\_parsing\_fails

doc = self.spacy\_nlp.parse(query)

File "spacy\_nlp.py", line 49, in parse

return SpacyDoc(self.\_\_instance.parser(query))

File "lib/python3.5/site-packages/spacy/language.py", line 328, in \_\_call\_\_

proc(doc)

File "spacy/syntax/parser.pyx", line 146, in spacy.syntax.parser.Parser.\_\_call\_\_ (spacy/syntax/parser.cpp:6114)

spacy.syntax.parser.ParserStateError: Error analysing doc -- no valid actions available. This should never happen, so please report the error on the issue tracker. Here's the thread to do so --- reopen it if it's closed:

https://github.com/spacy-io/spaCy/issues/429

Please include the text that the parser failed on, which is:

'splash On'

Here is our test:

nlp = spacy.en.English()

nlp.matcher.add('splash', 'my entity', {},[ [{LEMMA: 'splash'}, {LEMMA: 'on'}]])

nlp('splash On')

-------------------------------------------------------------------------

2017-02-01T22:08:54Z ejschoen

I'm afraid I'm getting this, too, in version 1.5.0:

```

File "nlp.py", line 166, in parse\_sentence

doc = nlp(sentence)

File "/usr/local/lib/python3.4/dist-packages/spacy/language.py", line 328, in \_\_call\_\_

proc(doc)

File "spacy/syntax/parser.pyx", line 146, in spacy.syntax.parser.Parser.\_\_call\_\_ (spacy/syntax/parser.cpp:6114)

spacy.syntax.parser.ParserStateError: Error analysing doc -- no valid actions available. This should never happen, so please report the error on the issue tracker. Here's the thread to do so --- reopen it if it's closed:

https://github.com/spacy-io/spaCy/issues/429

Please include the text that the parser failed on, which is:

'Located 16 km southwest of the Ceres find, the Hebe-1 well tested at 5,956 BOPD.'

```

All was fine, until I added some matcher rules and an on\_match callback:

```

def unit\_match\_cb(matcher,doc,i,matches):

spans = [(ent\_id, label, doc[start:end]) for ent\_id,label,start,end in matches]

for ent\_id, label, span in spans:

print(span, span[len(span)-1].tag)

span.merge(label=label, ent\_type='QUANTITY', tag='NNP')

def add\_uom\_match(matcher, unit):

matcher.add\_entity(unit, {"ent\_type": "QUANTITY"}, on\_match=unit\_match\_cb)

matcher.add\_pattern(unit, [{"like\_num":True},{ORTH: unit}], label="SPE\_UOM")

matcher.add\_pattern(unit, [{"like\_num":True},{ORTH: unit+"."}], label="SPE\_UOM")

matcher.add\_pattern(unit, [{"like\_num":True},{ORTH: unit+","}], label="SPE\_UOM")

matcher.add\_pattern(unit, [{"like\_num":True},{ORTH: unit+";"}], label="SPE\_UOM")

```

where unit is 'BOPD', for example. The on\_match callback is being called.

-------------------------------------------------------------------------

2017-04-30T05:03:37Z kyao

Got this error in version 1.7.3:

Traceback (most recent call last):

File "<stdin>", line 1, in <module>

File "/home/ktyao/anaconda3/envs/python27/lib/python2.7/site-packages/spacy/language.py", line 350, in \_\_call\_\_

proc(doc)

File "spacy/syntax/parser.pyx", line 207, in spacy.syntax.parser.Parser.\_\_call\_\_ (spacy/syntax/parser.cpp:7730)

spacy.syntax.parser.ParserStateError: Error analysing doc -- no valid actions available. This should never happen, so please report the error on the issue tracker. Here's the thread to do so --- reopen it if it's closed:

https://github.com/spacy-io/spaCy/issues/429

Please include the text that the parser failed on, which is:

u'Meet Linux.Mirai Trojan, a DDoS nightmare'

I am using a customized tokenizer that merges the three tokens, 'Linux', '.' and 'Mirai', into one token.

-------------------------------------------------------------------------

2017-06-22T06:38:49Z fsonntag

I'm also running in this issue on 1.8.2, nevertheless only after processing multiple documents in parallel:

```

File "spacy/syntax/parser.pyx", line 214, in spacy.syntax.parser.Parser.\_\_call\_\_ (spacy/syntax/parser.cpp:7989)

spacy.syntax.parser.ParserStateError: Error analysing doc -- no valid actions available. This should never happen, so please report the error on the issue tracker. Here's the thread to do so --- reopen it if it's closed:

https://github.com/spacy-io/spaCy/issues/429

Please include the text that the parser failed on, which is:

'NSPc1, a mainly nuclear localized protein of novel PcG family members, has a transcription repression activity related to its PKC phosphorylation site at S183. Nervous system polycomb 1 (NSPc1) shares high homology with verteb

rate PcG proteins Mel-18 and Bmi-1. The mRNA of NSPc1 is highly expressed in the developmental nervous system [Mech. Dev. 102 (2001) 219-222]. However, the functional characterization of NSPc1 protein is not clear. In the prese

nt study, using Western blotting technique, we aimed to describe the distributions of NSPc1 protein in rat tissues and cell lines. The subcellular localization of NSPc1 was examined in HeLa and SH-SY5Ycell lines, and its transc

riptional repression activity was examined in COS-7 cell line. We found that the NSPc1 protein was localized mainly in the nucleus. NSPc1 remarkably repressed the transcription. Most interestingly, both the C-terminal of NSPc1

and two phosphorylation sites in the C-terminal, especially the PKC phosphorylation site at S183, were important in mediating transcription repression. Taken together, results from our study suggest that NSPc1, as a typical PcG

family member, has powerful transcriptional repression ability, which may be related to the PKC signaling pathway.'

```

Edit: I think it's just the parallelization, that's not done by `nlp.pip`, but instead calling the parser from different threads. So nevermind :)

-------------------------------------------------------------------------

2018-05-08T19:28:13Z lock[bot]

This thread has been automatically locked since there has not been any recent activity after it was closed. Please open a new issue for related bugs.

-------------------------------------------------------------------------

------------------------------