Two examples of side-by-side, highlighted plagiarism display

1. Simple Melania Trump 2016 vs Michelle Obama 2008
http://www.huffingtonpost.com/entry/melania-trump-speech-plagiarized_us_578da752e4b0a0ae97c33675
http://www.news24.com/World/News/read-what-melania-trump-said-20160719

Melania Trump 2016 speech
"From a young age, my parents impressed on me the values
that you work hard for what you want in life, that your word is your
bond and you do what you say and keep your promise.
"That you treat people with respect

They taught and showed me values and morals in their daily life.
That is a lesson that I continue to pass along to our son, and we
need to pass those lessons on to the many generations to follow,
because we want our children in this nation
to know that the only limit to your achievements is the
strength of your dreams and your willingness to work for them.
I travelled the world while working hard in the incredible arena of
fashion.'   (*Sentence seems authentic Melania, not speechwriter...*)

This is a simple example of the display style, with just one
antecedent source and no reordering of words.
Two colors have been found sufficient for many cases:
Blue   Word-for-word, in order, text reformatted to match lines.
Yellow   trivial changes, doable by copy editor
~~Strikethru~~ for obvious simple deletions, unshaded if no match.

This is a subset of the more sophisticated scheme on next page,
which has been used for more complex cases.

Meredith Mciver (experienced) claimed that Melania read this
across the phone. Try reading without the two Barack and I. The
evidence is strong that (inexperienced) Melania did it, unsupervised.
http://www.nytimes.com/2016/07/20/us/politics/melania-trump-convention-speech.html
http://www.nytimes.com/2016/07/20/us/politics/melania-trump-speech.html

117 words, 66 blue (in-order, identical, counting "generation(s)")

Antecedent - Michelle Obama 2008
"Barack and I were raised with so many of the ~~same~~ values,
that you work hard for what you want in life, that your word is your
bond and you do what you say you're going to do;
that you treat people with ~~dignity and~~ respect, ~~even if you don't
know them, and even if you don't agree with them~~.
And Barack and I set out to build lives guided by these values,
and
pass them on to the next generation.
Because we want our children – ~~and all children~~ in this nation –
to know that the only limit to ~~the height of~~ your achievements is the
reach of your dreams and your willingness to work for them."

(Melania used "pass" twice, ambiguous which matches Michelle.)

This display style:
- Visually aligns the blue text, so is easy to verify.
- Offers a quick visual impression of the extent of direct copying.
Just as important, readers can then focus on the nature of:
- trivial edits, other changes, deletions
Readers might ponder the yellow words and strikethrus above.
For example, Obama wrote "values," whereas Trump wrote "lesson"
and "lessons." Nothing is wrong, but the wording is slightly odd,
sometimes a hint of plagiarism. "Reach" turned into "strength."

2. General approach

Blue	Word-for-word, in order, compared to direct antecedent, likely enough for simple cases. This includes singular/plural changes.

Yellow	trivial changes, doable by copy editor, sometimes done by plagiarists to lessen mechanical detection, but also proves not an innocent copy where someone intended to add quotes and forgot.

Green	Same text, rearranged order.

Orange	seems a close paraphrase, somewhat subjective.

Pink	Misrepresentation of a source, either cited, or copied.

Gray	oddity of some kind, or worth highlighting.

No color: might be original, or just not yet found.

To shade new text to match, one can select each of the 6 words above, then Shade>More Colors>OK, to put that into "Recent colors", easy to use at least from MS Word 2010 onward.

Getting rid of space-after-paragraph makes it easier to line up columns, and people rapidly come to see that the blue text is identical, and then can focus on the remainder.

My matching algorithm is ~manual approximation of UNIX *diff(1),* i.e., scan document, look for matches in antecedent(s), if necessary, skipping text moved, then come back and green shade moves.
Deep Climate (https://deepclimate.org/) used a slightly different match algorithm and each has plusses and minuses.
Mine tries to maximize blue text and usually does, but does not work so well for big moves of text.

I've sometimes had to use 3 columns to display plagiarism chains, where someone re-used text that was already plagiarized.

From experience, the most time-consuming work is *finding* the plagiarized antecedent text(s), Converting the following example took less than an hour to do.
Again, from experience, shading the blue and aligning them causes the other edits to become much more obvious, and often exposes mistakes where rewordings introduce errors.

This is an evolution of the styles seen in:
http://www.desmogblog.com/sites/beta.desmogblog.com/files/see.no_.evil_.speak_.little.truth_.pdf (2012)
All this started with *Deep Climate*'s use of fonts in 2009.
I tried using gray highlighting, and after some discussion we both switched to Cyan/Yellow highlighting. However, MS Word **Highlighting** uses intense colors and a limited palette. The current approach uses Word **Shading,** which offers a full color palette including muted colors.   Studying different shadings and display algorithms might be a good research problem for cognitive scientists.

Following is a more complex example of the application of this display style, with multiple antecedents and more editing.
It is more accessible than some of the more technical examples.

The original plagiarism was exposed by Ann Ribidoux here:
http://cabinetofplagiarism.blogspot.fr/2014/03/exhibit-old-school-theft_1.html	and I saw it via
http://andrewgelman.com/2014/05/14/plaig-publication-filter-together-last/ by Andrew Gelman
The following example just does the blue, (which is essentially algorithmic), and a few obvious yellows, and a few greens.
Those are slightly more subjective.
I didn't bother trying to do orange, as that is even more subjective, and when there is so much blue, it probably doesn't matter much.

3. Matthew C. Whitaker's Peace Be Still

3. Matthew C. Whitaker's Peace Be Still	**Various antecedents**

Peace Be Still (p.178 -- no citation)

Fueled by "angry white men" as well as by white women, an all-out battle for the life of the policy emerged. For Conservatives, the system was a zero-sum game that opened the door for jobs, promotions, or education to people of color while it shut the door on whites. In a nation that has celebrated the values of independence and "pulling oneself up by one bootstraps," conservatives soon argued that "unqualified" racial minorities were getting a "free ride" in American schools and in the workplace as a result of affirmative action policies. They referred to affirmative action incorrectly and contemptuously as a system of "preferential treatment" and "quotas." Some even claimed that many people of color enjoyed playing the role of "professional victim" to exploit the policy for their own benefit.

http://www.infoplease.com/spot/affirmative1.html

Fueled by "angry white men," a backlash against affirmative action began to mount. To conservatives, the system was a zero-sum game that opened the door for jobs, promotions, or education to minorities while it shut the door on whites. In a country that prized the values of self-reliance and pulling oneself up by one's bootstraps, conservatives resented the idea that some unqualified minorities were getting a free ride on the American system.

"Preferential treatment" and "quotas" became expressions of contempt. Even more contentious was the accusation that some minorities enjoyed playing the role of professional victim.

Peace Be Still (p. 179) no citation

...progressives pointed out that despite widespread fears that people of color were taking 'white people's jobs,' white men in particular still dominated the workforce by virtue of their positions, salaries, and prestige. The conflict over affirmative action was also a result of its ambiguous and complex evolution. Many Progressives, for instance, understood the injustices of the affirmative measures associated with *Wygant v Jackson Board of Education* (1986), in which black employees retained their jobs while white employees with seniority were laid off. In addition, many conservatives struggled to offer a better alternative to the imposition of a strict quota system in *United States* vs *Paradise*, et al (1987), in which the defiantly racist state officials of the Alabama Department of Public Safety refused to promote any African Americans above entry-level positions, even after they were mandated to do so by twelve years of court orders.

http://www.infoplease.com/spot/affirmative1.html

Liberals also pointed out that another popular conservative argument-that because of affirmative action, minorities were threatening the jobs of whites-belied the reality that white men were still the undisputed rulers of the roost when it came to salaries, positions, and prestige. The debate about affirmative action has also grown more murky and difficult as the public has come to appreciate its complexity. Many liberals, for example, can understand the injustice of affirmative action in a case like *Wygant* (1986): black employees kept their jobs while white employees with seniority were laid off. And many conservatives would be hard pressed to come up with a better alternative to the imposition of a strict quota system in *Paradise* (1987), in which the defiantly racist Alabama Department of Public Safety refused to promote any black above entry level even after a full 12 years of court orders demanded they did.

***Peace Be Still* (p.???) no citation**
Born Willard Christopher Smith Jr.
In Philadelphia, Pennsylvania, Smith attended Overbrook High School in
Winfield, Pennsylvania, where he soon became know
 as "The Prince" for his charm and quick wit.
  At the early age of
twelve, he began rapping and developing his own style under the influence
of hip-hop legend Grandmaster Flash. Just four years later, at only sixteen,
Smith met Jeff Townes, also known as DJ Jazzy Jeff, with whom he eventually
collaborated with (sic) under the title Fresh Prince. The two artists produced
a number of songs, including the worldwide hits "Girls Ain't Nothing But
Trouble" (1989) and "Parents Just Don't Understand" (1989), which won a
Grammy Award for Best Rap Performance

*Note the odd (sic), where "who … with" was copyedited to "with whom", but
the second "with" remained.*
*The correct name for the song is "Girls Ain't Nothing But Trouble" so in that
case, the text was improved.*

http://www.take40.com/artists/1272/will-smith/bio*
Will Smith was born Willard Christopher Smith Junior on September 25th,
1968 in Philadelphia, Pennsylvania. He attended Overbrook High School in
Winfield Pennsylvania, which is where he soon became known to his friends
as "The Prince," for his charming antics. Little did he know that this
nickname would still have meaning over 30 years later. At the early age of
12, Smith began rapping and developing his own style under the influence
of hip-hop legend, Grandmaster Flash. Four years later, at only 16,
he met Jeff Townes, also known as DJ Jazzy Jeff, who he eventually
collaborated with under the title, Fresh Prince. The two produced
a number of songs including the worldwide hit, Girls Ain't Nothin But
Trouble, and in 1989, the duo won their first Grammy for Best Rap
Performance for Parents Just Don't Understand.

* This has disappeared, but fortunately the Wayback Machine has it:
https://web.archive.org/web/20080905120037/http://www.take40.com/art
ists/1272/will-smith/bio

If the earlier file disappears, it has been architved as well, as at:
https://web.archive.org/web/20140227233100/http://www.infoplease.com
/spot/affirmative1.html